Dynamic Semiparametric Factor Models

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Doktor der Wirtschaftswissenschaft)
im Fach Statistik und Volkswirtschaft

eingereicht an der
Wirtschaftswissenschaftliche Fakultät
Humboldt-Universität zu Berlin

von
Herr Dipl.-Ing. M.Sc. Szymon Borak
geboren am 03.05.1979 in Breslau

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Dr. h.c. Christoph Markschies

Dekan der Wirtschaftswissenschaftliche Fakultät:
Prof. Ph. D. Oliver Günther

Gutachter:

1. Prof. Dr. Wolfgang Härdle
2. Prof. Dr. Christian Hafner

Tag des Kolloquiums: 9. July 2008

Abstract

High-dimensional regression problems which reveal dynamic behavior occur frequently in many different fields of science. The dynamics of the whole complex system is typically analyzed by time propagation of few number of factors, which are loaded with time invariant functions of exploratory variables. In this thesis we consider dynamic semiparametric factor model, which assumes nonparametric loading functions. We start with a short discussion of related statistical techniques and present the properties of the model. Additionally real data applications are discussed with particular focus on implied volatility dynamics and resulting factor hedging of barrier options.

Keywords:
dynamic semiparametric factor models, semiparametric regression, time series, asymptotic inference, forecasting, implied volatility, hedging, barrier option, electricity futures

Zusammenfassung

Hochdimensionale Regressionsprobleme, die sich dynamisch entwickeln, sind in zahlreichen Bereichen der Wissenschaft anzutreffen. Die Dynamik eines solchen komplexen Systems wird typischerweise mittels der Zeitreiheneigenschaften einer geringen Anzahl von Faktoren analysiert. Diese Faktoren wiederum sind mit zeitinvarianten Funktionen von explikativen Variablen bewichtet. Diese Doktorarbeit beschäftigt sich mit einem dynamischen semiparametrischen Faktormodell, dass nichtparametrische Bewichtungsfunktionen benutzt. Zu Beginn sollen kurz die wichtigsten statistischen Methoden diskutiert werden um dann auf die Eigenschaften des verwendeten Modells einzugehen. Im Anschluss folgt die Diskussion einiger Anwendungen des Modellrahmens auf verschiedene Datensätze. Besondere Aufmerksamkeit wird auf die Dynamik der so genannten Implizierten Volatilität und das daraus resultierende Faktor-Hedging von Barrier Optionen gerichtet.

Schlagwörter:
dynamische semiparametrische Faktormodelle, semiparametrische Regression, Zeitreihe, asymptotische Inferenz, Vorhersage, Implizierte Volatilität, Hedging, Barrier Optionen, Elektrizitätsfuture,

Acknowledgement

I would like to express deep gratitude to Professor Wolfgang Härdle for giving me the opportunity to come to Berlin and studying statistics and quantitative finance under his supervision. I highly appreciate that he has shown me the different perspective on quantitative modelling, which significantly improved my understanding of this topic.

I am thankful to Professor Christian Hafner for willingly accepting to evaluate my thesis and sit in the examination committee.

I would like to thank all those people with whom I collaborated during the preparation of this thesis. I am grateful to Professor Enno Mammen for inviting me to Mannheim and giving me valuable suggestions. The numerous comments, I have received from Professor Byeong Park, were for me essential for improving the simulation and estimation algorithms. My thanks go to Matthias Fengler for introducing me to the hedging problematic, Rafał Weron for explaining me electricity markets issues and Stefan Trück for close cooperation on CO_2 allowances prices application. I am grateful to Professor John-Dylan Haynes for supporting me with fMRI data and Jakob Heinzle for explaining me the fundamental issues of fMRI experiments.

I gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft by the Sonderforschungsbereich 649 Ökonomisches Risiko. I am greatly indebted to the staff of Financial and Economic Data Center for the technical support. During my work for the Institut für Statistik und Ökonometrie at Humboldt-Universität zu Berlin I benefited a lot from friendly and stimulating working environment created by my colleagues. I owe a special thanks to Michal Benko for his constant willingness for scientific discussions.

I highly appreciate a constant encouragement that I was receiving from my family and friends. Their support means a lot to me.

Berlin im Januar 2008 Szymon Borak

"A pessimist sees the difficulty in every opportunity; an optimist sees the opportunity in every difficulty."

Winston Churchill

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Problems worthy of attack prove their worth by fighting back."

Paul Erdos

Many physical, economical or biomedical phenomena provide repeated measurements of complex objects. In various cases the collected data reveal a nontrivial dynamic structure. Modelling for high-dimensional data is a challenging task in statistics and frequently requires a combination of flexible approach and dimension reduction methods. A prime example is the statistical analysis of financial term structure dynamics, which is crucial for bond portfolio risk management or derivative pricing, see Nelson and Siegel (1987) and Diebold and Li (2006). In financial engineering, it is common to analyze the dynamics of implied volatility surface for risk management, Cont and Fonseca (2002), Hafner (2004), Fengler et al. (2007), among others. In biomedical research the interest can be focussed on time improvement of the patients' treatment, as in the children growth history analysis by Martinussen and Scheike (2000). Other examples include the studies of radiation treatment of prostate cancer by Kauermann (2000) and evoked potentials in EEG analysis by Gasser et al. (1983). For functional magnetic resonance imaging data, one may be interested in analyzing the brain's response over time as well as identifying its activation area, see Worsley et al. (2002). In empirical macroeconomics one is interested in analyzing the dynamics of a plethora of economic indicators that reflect the state of the economy, see Stock and Watson (2005). In mortality analysis one builds a model for forecasting death rate or life expectancy, as is done in Lee and Carter (1992).

A successful modelling approach utilizes factor type models, which al-

low low-dimensional representation of the data. In an orthogonal L-factor model an observable J-dimensional random vector $Y_t = (Y_{t1}, \ldots, Y_{tJ})^\top$ can be represented as

$$Y_{tj} = m_{0j} + Z_{t1} m_{1j} + \cdots + Z_{tL} m_{Lj} + \varepsilon_{tj}, \tag{1.1}$$

where Z_{tl} are common factors, ε_{tj} are errors or specific factors and the coefficients m_{lj} are factor loadings. In most applications, the index $t = 1, \ldots, T$ reflects the time evolution of the whole system, and Y_t can be considered as a multi-dimensional time series. For a method to identify common factors in this model we refer to Peña and Box (1987). The study of high-dimensional Y_t is then simplified to the modelling of $Z_t = (Z_{t1}, \ldots, Z_{tL})^\top$, which is a more feasible task when $L \ll J$. The model (1.1) reduces to the generalized dynamic factor model considered by Forni et al. (2000), Forni and Lippi (2001), Hallin and Liska (2007), when $Z_{tl} = a_{l1}(B)U_{t1} + \ldots + a_{lq}(B)U_{tq}$, where the q-dimensional vector process $U_t = (U_{t1}, \ldots, U_{tq})^\top$ is an orthonormal white noise and B stands for the lag operator. In this case, the model (1.1) is expressed as

$$Y_{tj} = m_{0j} + \sum_{k=1}^{q} b_{kj}(B) U_{tk} + \varepsilon_{tj},$$

where $b_{kj}(B) = \sum_{l=1}^{L} a_{lk}(B) m_{lj}$.

In a variety of applications, one has explanatory variables X_{tj} at hand that may influence the factor loadings m_l. An important refinement of the model (1.1) is to incorporate the existence of observable covariates X_{tj}. The factor loadings are now generalized to functions of X_{tj}, so that the model (1.1) is generalized to

$$Y_{tj} = m_0(X_{tj}) + \sum_{l=1}^{L} Z_{tl}\ m_l(X_{tj}) + \varepsilon_{tj}. \tag{1.2}$$

In particular, lagged observations of the component series $(Y_{t-1,j}, \ldots, Y_{t-p,j})^\top$ may be the regressor X_{tj}, but we allow here any external variable $X_t = (X_{t1}, \ldots, X_{tJ})^\top$. Additionally, the regularity of the multi-dimensional time series may be omitted by allowing J depending on t, say J_t. This structure may resemble the longitudinal data when the data reflects a measurements of particular subjects through the time. This condition, however, is not directly incorporated in (1.2).

The model (1.2) assumes an additive structure, which is common in statistical modelling. This framework allows for an easy separation between

spatial and temporal components, which significantly simplifies the inference. In fact, it can be interpreted as a discretized version of the following functional extension of the model (1.1)

$$Y_t(x) = m_0(x) + \sum_{l=1}^{L} Z_{tl}\ m_l(x) + \varepsilon_t(x), \tag{1.3}$$

where $\varepsilon_t(\cdot)$ is a mean zero stochastic process, and also regarded as a regression model with embedded time evolution. The functional form of (1.3) has strong relations to functional data analysis, see Ramsay and Silverman (1997). In particular the factor form could be obtained by applying the functional principal components analysis.

Treating (1.2) as a regression model provide as well slightly different perspective. Were Z_t be observable the model would reduce to the varying-coefficient model, such as in Fan et al. (2003) and Yang et al. (2006). On the other hand, neglecting index t one obtains regression based on a pooled data. Cancelling Z_t and assuming that $m_l(\cdot)$ are nonparametric functions of one coordinate yields semiparametric additive model, see Hastie and Tibshirani (1990). If the model is again enhanced by including time dependent component this setting is similar to the one considered in Connor and Linton (2007) and Connor et al. (2007), which generalized the study of Fama and French (1992) on the common movements of stock price returns.

In this thesis we consider the model (1.2) with nonparametric functions m_l and unknown Z_t. We call this model *dynamic semiparametric factor model* (DSFM). The estimates $\widehat{m}_l$ and $\widehat{Z}_t$ are obtained by direct discrete representation applied to (1.2) and minimizing a least square loss. On the one hand, it is nothing else but the nonparametric extension to models with parametric functions like Nelson and Siegel (1987) for yield curve or Hafner (2004) for implied volatility. On the other hand, it is a new nonparametric modelling perspective. It is different from functional principal components since the functions comes not from spectral decomposition of the functional covariance operator. It is also quite different from Connor and Linton (2007), where the covariates, denoted there by $X_{l,j}$, are time-invariant and are different for different m_l, which allows a direct application of backfitting procedures.

Some ambiguities may arise from the name of the model. The word *factor* have clear justification by relation to the traditional orthogonal factor model. However, some confusion may come from the *dynamic factor model*. The model of Geweke (1977) is popular in time series literature and the common factors are incorporated together with its lags. Here one would rather link word *dynamic* and *semiparametric* to indicate time changing semiparametric regression or *semiparametric* and *factor* to underline the estimation method.

The factors stay 'static' and it seems to be justifiable to split words *dynamic* and *factor* with term *semiparametric*. Studying the case where the lags of the factors are included goes far beyond the scope of this thesis.

It has to be mentioned that the DSFM was proposed by Fengler et al. (2007) and one of its main motivation comes from the special structure of the implied volatility data. For more details on the data design and economic importance we refer to Section 4.1.1. Fengler et al. (2007) fit the functions $\widehat{m}_l$ only in the local neighborhood of the design points in order to avoid possible model bias or initial implied volatility surface estimation. Some properties of their algorithm is studied in Borak et al. (2005). Brüggemann, Härdle, et al. (2006) perform an econometric analysis to the estimated factors and Giacomini and Härdle (2007) applied the resulting estimates to analyze the empirical pricing kernels. The method produces estimates of the true unobservable Z_t, say $\widehat{Z}_t$, as well as estimates of the unknown functions m_l. In practice, one operates on these estimated values of Z_t for further statistical analysis of the data. Borak et al. (2007) show that for any version of Z_t there exists a version of $\widehat{Z}_t$ whose lagged covariances are asymptotically the same as those of Z_t. This justifies the inference based on $\widehat{Z}_t$ when Z_t, in particular is a VAR process.

The main goal of this thesis is to present and illustrate the properties of the DSFM. We perform on three different platforms. First, we show how the model is related to other statistical techniques; these are presented only in a basic manner. The aim is to give rather an intuitive exposition and omit technical details that have less relevance to the main method. Some technical results are only provided for the DSFM itself.

Second, we try to approach the complexity of the model by simulation studies. Some of the numerical experiments illustrates the theoretical results, for example the asymptotical equivalence of the inference on the true and estimated time series of factors. Other experiments show the finite sample behavior of the model, which is hard to deduce from theoretical studies only. For the simulation studies only one particular setting is chosen, which preclude easy generalization of the obtained result. However, we believe that our setting is representative and nontrivial. Due to the complexity of the structure and computation limitations more extended simulation analysis seems to be infeasible in the moment.

Third, our illustration focuses on the real data performance. The factor model (1.2), in particular the DSFM, has numerous applications. A prominent example is modelling of yield curve evolution. The standard approach is to use the parametric factor model proposed by Nelson and Siegel (1987), where the empirical form of the yield curve is fitted with some pre-specified functions of the bonds' maturities. A possible refinement is the penalized

spline smoothing, employed simultaneously in time and maturity dimensions, see Krivobokova et al. (2006). Apart from this, other modelling strategies based on principal component analysis or factor models are commonly used, see Rebonato (1998), Bliss (1997) and Molgedey and Galic (2001) among others. Clearly, the DSFM would enhance flexibility of these approaches for modelling the term structure of interest rates. In a similar manner the DSFM may be employed to analyze the term structure of the variance swaps as in Detlefsen and Härdle (2006). Other example is mortality trend fitting, where the current standard is to use a model proposed by Lee and Carter (1992). In that model the age-specific death rates are regressed additively on a time-invariant age-specific component and another age-specific component multiplied by a time-varying factor. Here, one can let the age specific components be nonparametric functions of some particular covariates and extend the model to the reduced DSFM with $L = 1$.

In this thesis, however, we focus on the application to CO_2 emission allowances prices, functional magnetic resonance imaging (fMRI) data, electricity futures prices and 'traditional' implied volatility surface dynamics. The model fitted to implied volatility dynamics can be seen as a low-dimensional projection of the implied volatility risk factors. Certain exotic options, whose prices strongly depend on the whole surface, have to be hedged additionally against undesirable surface movements. This, so called *vega hedging*, can be refined by adjusting the hedging portfolio according to the new hedging ratios, defined as the derivatives of the option price with respect to Z_t. We concentrate on the empirical performance of this procedure applied to the barrier options when the factors are extracted through the DSFM. As the model provides a useful tool for the curve forecasting its prediction power is evaluated on the electricity futures prices term structure. The DSFM is confronted with a principal component approach, which gives a comparison of two modelling perspectives.

The thesis is organized as follows. Chapter 2 gives an overview over models and statistical techniques that are related to the DSFM. We keep our presentation short and show only some basic notions. First, an idea of series approximation is given since it has crucial meaning in presenting the estimation algorithm. Then, the standard statistical definitions are recalled and regression problematic is sketched. The next idea that we find useful to illustrate is time series modelling since one of the main aims of the DSFM is to simplify complex dynamics to the low-dimensional time series modelling of factors. Traditional time series data can be extended to the panel data when some external variable is at hand and for completeness basic methodology concerning this problematic is provided. In the next step functional data analysis is discussed and functional principal component analysis is shown,

since this methodology has very close relation to the DSFM. We close the excursion on the statistical approaches by presenting parametric, functional data analysis based and semiparametric factor models.

The last section of Chapter 2 concentrates fully on the presenting the DSFM. We start by some graphical recapitulation of the statistical techniques, which lead to the DSFM. Then we present three possible estimation methods and give some asymptotic results. Finally, issues concerning model choice and model identification are discussed.

Chapter 3 focuses on simulation studies. We start by presenting the setting, from which the artificial data are generated. We fix time series model for Z_t and the form of the functions m_l, and analyze estimated $\widehat{Z}_t$ and $\widehat{m}_l$. Our conclusions are based on the several repetitions. By presenting the distributional behavior of the covariance matrices' differences we illustrate the asymptotic theorem from Chapter 2. By choosing different noise level, number of observations per time point, time series length and spline number we demonstrate how sensitive is the model and estimation algorithm with respect to these quantities. The forecasting experiment is performed and finally we show how the data design may influence the estimation results.

Chapter 4 presents the applications of the model. We concentrate on four types of problems: dynamics of implied volatility surfaces, dynamics of CO_2 allowances futures prices, electricity futures term structure and application to functional magnetic resonance imaging. In Section 4.1 the corresponding problematic is introduced, references to traditional methodology are provided and the detailed descriptions of the data sets are given. Section 4.2 shows how the DSFM can be fitted to the presented data sets. In the next sections we evaluate the fitted model. In Section 4.3 the forecasting power is studied on the electricity future term structure in a moving window framework. Section 4.4 concentrates on a hedging experiment. The model serves as a tool for identifying nonstandard risk factors and empirical dynamic hedging procedure for reverse barrier options is studied in local volatility framework.

Chapter 2

Theoretical Background

"As far as the laws of mathematics refer to reality, they are not certain, as far as they are certain, they do not refer to reality."

Albert Einstein

The progress in science is almost ever based on minor improvements of the previous achievements. Similarly in the statistical modelling new ideas comes often as improvements of the traditional ones or as a merging of several different techniques. Therefore, in order to have better understanding of the statistical method, it is worth to recognize it as a generalization or special case of other procedures.

The aim of this chapter is to present the theoretical background of the dynamic semiparametric factor model. Related statistical issues are sketched in order to embed the model in the traditional statistical methodology. Additionally several properties are discussed. The exposition starts with presenting series representation for a smooth function. Then Section 2.2 gives the basic notions in statistics and Section 2.3 concentrates on the regression problem. Factor models are discussed in Section 2.4 and time series issues in Section 2.5. Some basic notions of panel data and functional data analysis are presented in Section 2.6 and Section 2.7 respectively. The dynamic regression problems are shown in Section 2.8 and the last section fully concentrates on the dynamic semiparametric factor model.

The goal of this chapter is to give only very short illustration on the described methods. However, the problematic that we discuss in each section requires a separate monograph for comprehensive presentation. We believe that by presenting only few issues in very limited extension, the at-

tention would be focussed to the main method and the detailed exposition can be found in positions which fully concentrate on each particular problematic. Just to give few examples splines problematic is presented in Boor (2001), nonparametric regression is discussed in Härdle (1990), time series in Lütkepohl (2004) and functional data analysis in Ramsay and Silverman (1997). The technical proofs of Section 2.9 can be found in the original paper of Borak et al. (2007).

2.1 Series Representation

Before we present the statistical issues related to the data analysis, we discuss briefly some basic ideas from the pure mathematical analysis and mathematical engineering. The main emphasis is on the fundamental concepts of function representation, which are essential for our later studies. We find presenting this idea important since the further discussion of handling the data in functional sense have the roots in finite function representation.

Consider some normed function space $\mathbb{B}$ over algebra $\mathbb{R}$ with infinite but countable basis (ψ_n). Then any function $f \in \mathbb{B}$ can be written as

$$f = \sum_{i=1}^{\infty} a_i \psi_i,$$

where $a_n \in \mathbb{R}$. This result is of the main importance for the approximation of the function f. In fact replacing the infinite sum with first K elements yields immediately to the convenient representation of f. Although some information about the function is lost but one obtains parsimonious structure, vector of the length K, which allows efficient handling of the complex functional object. Obviously, the quality of the approximation depends on K and (ψ_n). There is a plethora of the possibilities for the choice of (ψ_n) and below some of them are introduced.

Certainly, the idea presented above is not the only possible discrete representation. Another common way is to obtain the values of the function on some discrete set of arguments, e.g. regular grid, and define the approximation/extrapolation procedure of mapping from this particular set to any argument, e.g. linear approximation.

Taylor Expansion

One of the basic results from the mathematical analysis is a possibility of expanding a function in a local neighborhood using some polynomial terms

and the function's derivatives. Under certain regularity assumptions a one-dimensional Taylor series expansion of a real function $f(x)$ about a point $x = a$ is given by

$$f(x) = f(a) + \sum_{i=1}^{+\infty} \frac{f^{(i)}(a)}{i!}(x-a)^i.$$

This formula automatically suggests an idea for the finite approximation, by exchanging the infinity sum with finite sum. This motivates the construction of polynomial basis by taking first p polynomials of form $(x-a)^i$. In general d-dimensional case the Taylor series has a form of

$$f(x) = \sum_{i_1=0}^{+\infty} \dots \sum_{i_d=0}^{+\infty} \frac{\partial^{i_1}}{\partial x^{(1)^{i_1}}} \dots \frac{\partial^{i_d}}{\partial x^{(d)^{i_d}}} \frac{f(a)}{i_1! \dots i_d!} (x^{(1)} - a^{(1)})^{i_1} \dots (x^{(d)} - a^{(d)})^{i_d}$$

where $a = (a^{(1)}, \dots, a^{(d)})^\top$ and $x = (x^{(1)}, \dots, x^{(d)})^\top$.

Splines

For the general case of smooth functions the presented idea of Taylor series approximation performs relatively well only in the local neighborhood of point a. For the arguments far from a the small order polynomial might yield unsatisfactory results. To overcome this hurdle the standard approach is to build the polynomial approximation on many points. For a given set of points $x_1, \dots, x_p$, known as *breakpoints* or *knots*, one builds several polynomials and cuts them such that the polynomial expression is valid only locally. The knots play the role of binding points which might guarantee certain smoothness of the resulting function. This piecewise polynomial construction is known as *splines*.

To be more specific consider $d = 1$, a compact interval $[a, b]$ and a sequence of knots $a = x_0 < \dots < x_K = b$. Let $\varphi_1, \dots, \varphi_K$ be polynomials of order p. Then the spline is defined as

$$\begin{aligned} s(x) = \varphi_1(x) \quad &\text{for} \quad x \in [x_0, x_1) \\ s(x) = \varphi_2(x) \quad &\text{for} \quad x \in [x_1, x_2) \\ &\vdots \\ s(x) = \varphi_K(x) \quad &\text{for} \quad x \in [x_{K-1}, x_K]. \end{aligned}$$

Since the polynomials have p constants there are certain degrees of freedom for the splines' construction. It is common to impose some smoothness conditions to use splines efficiently in applications. For a detailed discussion on

the splines's problematic (how can different spline families be defined and the overview of their properties) refer to Boor (2001). Here we present only the idea of B-spline basis which is further utilized in our discussion.

B-splines are very popular functional basis, which offers flexibility and relatively convenient and easy evaluation of the derivatives. They can be defined by recursive algorithm. For a given sequence of knots let $B_{k,p}(x)$ denote a kth B-spline of order p. The B-spline of order $p = 1$ is given by an indicator

$$B_{k,1}(x) = \begin{cases} 1 & \text{if} \quad x \in [x_{k-1}, x_k) \\ 0 & \text{otherwise.} \end{cases}$$

Let for completeness $B_{0,p}(x) = 0$ and $B_{K+p,p}(x) = 0$. Set additionally $p-1$ multiple knots at the end points of the interval namely $x_{-p+1} = \ldots = x_{-1} = x_0$ and $x_K = x_{K+1} \ldots = x_{K+p-1}$. Then the higher order B-splines are defined with lower order ones by recursive formula

$$B_{k,p}(x) = \frac{x - x_{k-p}}{x_{k-1} - x_{k-p}} B_{k-1,p-1}(x) + \frac{x_k - x}{x_k - x_{k-p+1}} B_{k,p-1}(x),$$

where $k = 1, \ldots, K + p - 1$. The resulting indeterminate term of type $\frac{0}{0}$ we set to 0. Note that for $K + 1$ knots one obtains $K + p - 1$ B-splines.

For the basis illustration how the splines are constructed we refer to Figure 2.1. We take $K = 4$ for equidistance knots and plot linear, quadratic and cubic B-splines basis. The initial constant basis would just be one between the knots and zero otherwise and we omit to present it here. The pth order B-splines are usually nonzero on p intervals inferred by knots' sequence. This is not the case for the boundary splines which are nonzero only on $1, 2, ..., p-1$ intervals. This feature is caused by assuming multiple knots at the ends of the interval. The boundary behavior of this basis may result in some unwanted explosion of the approximated function.

Polynomial Based Basis

We have already discussed the classical polynomial basis motivated by Taylor expansion and the B-spline basis which are piecewise polynomials. Other possible choice is to consider truncated polynomial basis. The basis functions have the form

$$1, x, \ldots, x^K, (x - x_0)_+^K, \ldots, (x - x_K)_+^K.$$

First $K + 1$ are polynomial up to degree K and others are truncated power functions of degree K. The truncating functions can be also seen as building blocks of B-spline.

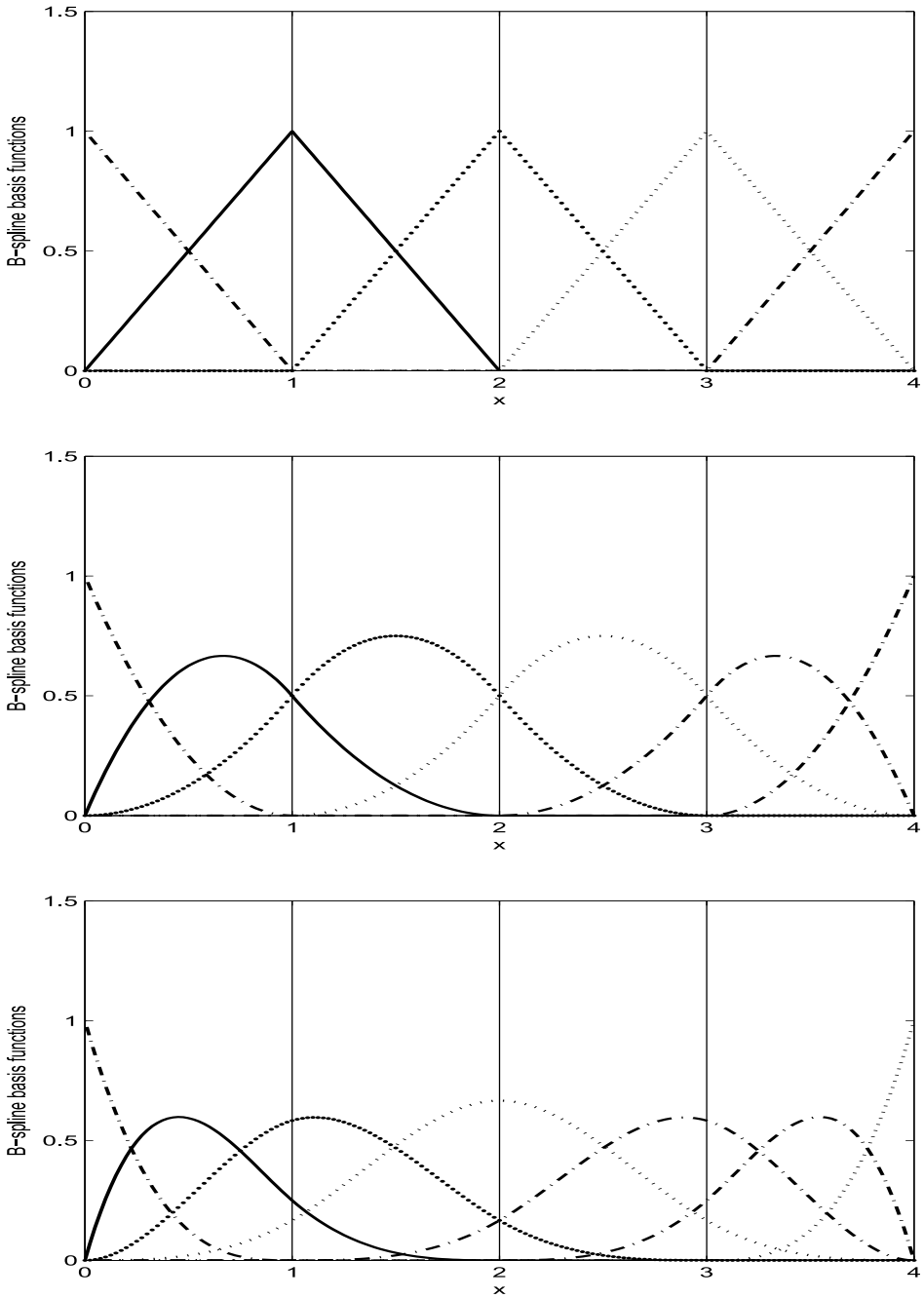

Figure 2.1: *Example of one dimensional B-splines. Upper panel: linear splines. Middle panel: quadratic splines. Lower panel: cubic splines. The solid vertical lines indicate the knots' placement.*

The other polynomial based basis are Chebyshev polynomials of the first kind

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x), \end{aligned}$$

the Chebyshev polynomials of the second kind

$$\begin{aligned} U_0(x) &= 1, \\ U_1(x) &= 2x, \\ U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x), \end{aligned}$$

Legendre polynomials

$$L_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} \left\{ (x^2 - 1)^n \right\},$$

or Hermite polynomials

$$H_n(x) = (-1)^n \exp(x^2/2) \frac{d^n}{dx^n} \exp(-x^2/2)$$

Contrary to the B-spline basis these are the orthogonal series families. However, they are not equal to zero on some compact intervals and would reflect rather global features while B-splines reflects the local features by its piecewise character.

Fourier Series

A very popular basis commonly used in many engineering fields are provided by Fourier series. The basis functions have the following form

$$\begin{aligned} \psi_0(x) &= 1, \\ \psi_{2r-1}(x) &= \sin r\omega x, \\ \psi_{2r}(x) &= \cos r\omega x, \end{aligned}$$

for $r = 1, 2, \ldots$. They are orthogonal and periodic with the period determined by the parameter ω. Therefore it could be beneficial to use this basis for modelling periodic events like weather or electricity loads.

2.2 Statistical Modelling

The previous discussion focused on the concepts from mathematical analysis. Here we give very brief introduction to some basic notions from mathematical statistics. The more detailed explanations can be found in any textbook on probability and statistical theory, e.g. Shao (2003), Billingsley (1995), Feller (1968) etc.

The main interest of statistics is in extracting information from some *data*. The data are defined as a result of some series of random experiments. A *random variable* X is a measurable function from some probability space $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}^d, \mathcal{B}^d)$, where $\mathcal{B}^d$ is a borel σ-field. The *data set* or *sample* is then a collection of elements $(x_1, ..., x_n)^\top$, where $x_i \in \mathbb{R}^{d_i}$ is an outcome of some random variable X_i.

A *statistical model* is a set of assumptions on probability measure P. The main aim of statistics is to infer P from sample $(x_1, ..., x_n)^\top$ in order to make an inference on $(\Omega, \mathcal{F}, P)$.

Let $\mathcal{P}_\theta$ be a set of probability measures on $(\Omega, \mathcal{F})$. If the indexing parameter θ is taken from some finite dimensional subset $\Theta \subset \mathbb{R}^d$, then the family $\{\mathcal{P}_\theta\}$ is called *parametric family. Nonparametric family* is a family of probability measures which is not parametric. Similarly the models where the probability measure P comes from parametric (nonparametric) families are called parametric (nonparametric) models.

The parametric models are much easier to handle since the finite dimensional parameter specify the whole distribution. On the other hand nonparametric models allows more flexible modelling since they are not squeezed by finite dimensional assumption. In order to achieve some compromise between tractability and flexibility so called semiparametric models can be considered. They keep infinite dimensional parameter but impose some restriction on the parameter space. A standard example is decomposing θ into parametric and nonparametric parts, i.e. $\theta = (\theta_1, \theta_2)^\top$ where θ_1 is finite dimensional and θ_2 is infinite dimensional. Some discussion on properties of parametric, nonparametric and semiparametric methods can be found in Härdle et al. (2004).

2.3 Regression

Regression models are of the great importance for modelling non-i.i.d. data. The main goal is to establish a systematic relationship between two random variables X and Y. In the classical formulation one observes the data in form of $(X_1, Y_1), \ldots, (X_T, Y_T)$. From now on we do not make a distinction between the random variable and its realization, and what is meant will be

clear from the context. We assume in this section $X_t \in \mathbb{R}^d$ and $Y_t \in \mathbb{R}$ for $t = 1, \ldots, T$. X_t are called *explanatory* variables since they explain the systematic behavior of Y_t *dependent* variables.

The regression model is expressed as

$$Y_t = f_\theta(X_t) + \varepsilon_t, \tag{2.1}$$

where $f_\theta : \mathbb{R}^d \to \mathbb{R}$ is the regression function describing the systematic relationship. ε_t denotes the observation's specific component and explain the residual part not included in the regression function. The common assumption states that ε_t is a random error with $E\varepsilon_t = 0$. Therefore the model (2.1) can be equivalently written

$$E(Y_t|X_t) = f_\theta(X_t). \tag{2.2}$$

One of the estimation issues in (2.1) and (2.2) is to specify the form of function f_θ. It is essential in building some inference on $Y_{t'}$ when only a corresponding $X_{t'}$ is at hand. One possible approach is to search for such a function $\widehat{f}_\theta$ that minimizes empirical square residuals, i.e. $\sum_{t=1}^{T} \{Y_t - f_\theta(X_t)\}^2$. This loss function could be refined in many possible directions by considering e.g. absolute error, median or a combination of quadratic and absolute losses. Assuming some distributional form for ε_t e.g. i.i.d. Gaussian, would lead to the another set of powerful estimation methods. Obviously one searches within some specified class of functions and some examples are given below.

2.3.1 Parametric Regression

As a first possible choice for the function f_θ we discuss briefly the case when the function is fully specified with the finite dimensional parameter. A very popular model is linear regression, where the function is a linear combination of parameters and variable coordinates. The model can be written

$$E(Y_t|X_t) = Z_0 + \sum_{j=1}^{d} Z_j X_t^{(j)}, \tag{2.3}$$

where $X_t^{(j)}$ denotes jth coordinate of X_t. The parameter $\theta = (Z_0, ..., Z_d)^\top$ is a finite dimensional parameter and fully specify whole regression function.

In order to estimate the parameters consider a matrix notation for (2.3)

$$Y = X^\top Z + \varepsilon,$$

where $Y = (Y_1, \ldots, Y_T)^\top$. X is now a matrix with tth row set to $(1, X_t^\top)$ and ε denotes the vector of errors. Then the least square estimator is given by

$$\widehat{Z} = (XX^\top)^{-1}XY.$$

Implementing some weights, motivated for example by heteroscedastic errors, leads to minimizing $\sum_{t=1}^{T} w_t \left\{Y_t - Z_0 - \sum_{j=1}^{d} Z_j X_t^{(j)})\right\}^2$. The estimator is then refined to

$$\widehat{Z} = (XWX^\top)^{-1}XWY. \tag{2.4}$$

where W is a diagonal matrix of weight.

The model (2.3) can be refined as well by including some known functions $g_j : \mathbb{R} \to \mathbb{R}$ for $j = 1, ..., d$. Then it has a form

$$E(Y_t|X_t) = Z_0 + \sum_{j=1}^{d} Z_j g_j(X_t^{(j)}). \tag{2.5}$$

An example $g_j(x) = x^j$. If one additionally impose the restriction on X_t such that $X_t^{(1)} = ... = X_t^{(d)}$ then the model (2.5) is extended to the case of polynomial regression. Obviously there are a plethora of possible parametric regression function settings that we do not consider here. We only briefly introduced some ideas that have a direct consequence on the further discussion.

2.3.2 Nonparametric Regression

In parametric regression case one assumes that the function is specified only by finite dimensional parameter. However, the assumed structure, e.g. linear, in many situations could be too restrictive and leads to misleading results. Therefore a nonparametric regression is often considered, where the function f_θ cannot be fully specified by the finite dimensional parameter. Obviously some smoothness assumption should be made on the function f_θ since otherwise it would be impossible to distinguish between random and systematic component.

Although f_θ does not come from parametric family for practical usage it has to be approximated by some finite representation, see Section 2.1. However, if this approximation does not comes from the modelling condition but from practical needs one still remains in the nonparametric framework.

As an example we present a well known kernel Nadaraya-Watson estimator. The regression function estimator is given by

$$\widehat{f}(x) = \frac{\sum_{t=1}^{T} \mathcal{K}_h(x - X_t)Y_t}{\sum_{t=1}^{T} \mathcal{K}_h(x - X_t)}. \tag{2.6}$$

The kernel function $\mathcal{K}_h(\cdot)$ plays a role of weighting of the particular point X_t for each argument x. It is common to take symmetric probability density function for $\mathcal{K}_h(\cdot)$. In case $d > 1$ we will simplify the analysis by taking product kernels. The scale parameter h associated with the kernel function, called bandwidth, corresponds to the size of the weight. For the product kernel h is a d-dimensional vector. While the choice of bandwidth may have strong consequences for the shape of the estimator the choice of kernel is typically of minor importance, see Härdle (1990).

A natural refinement of the Nadaraya-Watson estimator is local polynomial estimator. Let for the moment $d = 1$ and the tuple of functions $\mathrm{f} = (f, \ldots, f^{(p)})^\top(x)$ denotes function f and all derivatives up to order p. Then taking $p + 1$ terms of Taylor series, see Section 2.1, and evaluating weighted least squares criterion with kernel weights leads to estimator

$$\widehat{\mathrm{f}}(x) \quad = \quad \arg\min_{\mathrm{f}(x)} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \frac{f^{(j)}(x)}{j!}(X_i - x)^j \right\}^2 \mathcal{K}_h(x \; - \; X_i). \quad (2.7)$$

Note that local polynomial estimator yields not only values of the function but also the values of its derivatives. The Nadaraya-Watson estimator can be regarded as a special case of local polynomial estimator for $p = 0$. For each particular point x the estimation algorithm (2.4) maybe adapted. The case of $d > 1$ is straightforward but tedious. The multivariate local linear estimator is discussed in Härdle et al. (2004) or Fan and Gijbels (1996).

The unknown function f can be accordingly represented by a weighted sum of a basis function like in Section 2.1. Then the finite approximation scheme allows to proceed like in parametric regression by regressing Y_t on $\psi_1(X_t), \ldots, \psi_K(X_t)$. The estimates $(\widehat{a})_{j=1,\ldots,K}$ combined with basis functions $(\psi)_{j=1,\ldots,K}$ yield the estimate of θ. Note that while K is selected one proceed like in parametric models, but the freedom for the choice of K keeps this method in nonparametric framework.

2.3.3 Semiparametric Regression

Semiparametric regression is when some structure is imposed but still the function f_θ cannot be represented only by the finite dimensional parameter. The functional form is not directly assumed however the structure of the model leaves less flexibility than in the nonparametric case. One of the motivation for creating this limitation, comes from the curse of dimensionality, since in high dimensions the nonparametric methods may become infeasible.

Among many possible semiparametric models, see Härdle et al. (2004) or Horowitz (1998) for more comprehensive overview, we focus here on the illustration based on the additive model. The key assumption is that the regression function has an additive structure of one dimensional functions of the explanatory variable coordinates. The model can be then written as

$$E(Y_t|X_t) = \sum_{j=1}^{d} g_j(X_t^{(j)}), \tag{2.8}$$

where $g_j : \mathbb{R} \to \mathbb{R}$ are unknown functions. Had the functions specific form e.g. linear one would obtain the classical parametric regression. Were there no additive structure the model would correspond to the pure nonparametric regression. The form of (2.8), however, place it somewhere between, namely in semiparametric regression.

The assumption of one dimensional functions is not exclusive. However, we would like to present the model as a extension of the classic linear regression. By such a formulation one simplifies the analysis by neglecting the interactions. In (2.8) we also omit the intercept term. It is due to the fact that the model can be identified only up to the shifts transformations so the implicit assumption is that intercept, $E\{g_j(X_t^{(j)})\}$ and mean of Y_t are all zero.

The model (2.8) is commonly estimated with a backfitting algorithm. It is an iterative procedure which consecutively updates the component functions. The main idea comes from the theory of projection in Hilbert spaces. The best additive function is understood as the closest to nonrestricted function, where the distance is meant as a expected squared distance. As a consequence

$$E\left\{Y_t - \sum_{j \neq j'} g_j(X_t^{(j)})|X_t^{(j')}\right\} = 0, \tag{2.9}$$

and

$$g_{j'}(X_t^{(j')}) = E\left\{Y_t - \sum_{j \neq j'} g_j(X_t^{(j)})|X_t^{(j')}\right\}. \tag{2.10}$$

The estimators from Section 2.3.2 can be plugged in in case all other functions are known. This motivates the following iterative algorithm:
(i) Initialize estimetes $g_1^{(0)}, \ldots, g_d^{(0)}$.
(ii) Repeat the cycles for $j' = 1 \ldots, d$ and estimate $\widehat{g}_{j'}^{(i+1)}$ with some nonparametric method applied to (2.10), where the last available estimator is plugged in for g_j $(j \neq j')$
(iii) Stop when the functions do not change significantly.

For the more comprehensive overview we refer to Hastie and Tibshirani (1990), where the idea of the backfitting is motivated in more details. In fact from (2.9) one may obtain system of $(Td \times Td)$ equations and the presented iterative procedure is Gauss-Seidel method for obtaining the solution. Note that the method depends on the initial estimates, which can be taken as linear regression estimates or simply zero. Note also that ordering the variables may have influence on the final estimates as well. In case of $d = 2$ one does not need iterative algorithm and the backfitting estimator can be found explicitly, see Härdle et al. (2004). Different methods for estimating the additive model are some evolution of classical backfitting. The typical example are modified backfitting like in Hastie and Tibshirani (1990), smooth backfitting from Mammen et al. (1999) or marginal integration as shown in Linton and Nielsen (1995).

2.4 Factor Models

This section gives a brief overview of factor analysis. The main motivation of this statistical technique comes from the necessity of dimension reduction. Suppose that variables are clustered according to their correlation. It could be reasonable instead of studying the observation in its full dimensionality to reduce the analysis to the small number of meaningful factors. If this kind of simplification is subjected only to the negligible loss of information, it gives valuable insight to the core of the problem. Here we present some issues concerning factor analysis in classical multivariate framework. For more details we refer to Johnson and Wichern (1998), or Härdle and Simar (2003).

Let $Y_t \in \mathbb{R}^J$ for $t = 1, \ldots, T$. For the simplicity of notation we assume $\sum_{t=1}^{T} Y_t = 0_J$, where 0_J is a vector of zeros. If it is not the case one may always consider mean adjusted vector.

The factor model has the following form for $j = 1, \ldots, J$ and $t = 1, \ldots, T$

$$Y_{jt} = m_{j1} Z_{1t} + \ldots + m_{jL} Z_{Lt} + \varepsilon_{jt}, \tag{2.11}$$

where m_{jl} are called factor loadings, Z_{lt} are common factors for $l = 1, \ldots, L$ and ε_{jt} are called specific factors. Contrary to regression problems neither m_{jl} nor Z_{lt} are observable.

Rewriting (2.11) into a matrix notation gives

$$Y_t = M Z_t + \varepsilon_t, \tag{2.12}$$

where M is a $(J \times L)$ matrix of loadings (m_{jl}), $Z_t = (Z_{1t}, \ldots, Z_{Lt})^\top$ is a $(L \times 1)$ vector of factors and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Jt})^\top$ is a $(J \times 1)$ vector of specific

factors. $Y_t = (Y_{1t}, \ldots, Y_{Jt})^\top$ is a $(J \times 1)$ vector of observations. In more compact matrix notation the model has a form

$$Y = MZ + \varepsilon, \tag{2.13}$$

where $Y = (Y_1, \ldots, Y_T)$ is $(J \times T)$ matrix, $Z = (Z_1, \ldots, Z_T)$ is $(L \times T)$ matrix and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)$ is $(J \times T)$ matrix.

With standard assumptions that ε_t is a zero mean random error, the statistical analysis of Y_t may be conveniently replaced by the statistical analysis of Z_t when $L \ll J$. Inference on Z_t can be transferred to the inference on Y_t through a matrix of loadings, since it is common for all observations. Therefore, one obtain a powerful method that may significantly simplify the analysis on the complex system.

2.4.1 Identification

The model (2.13) cannot be uniquely identified. For any invertible $(L \times L)$ matrix B it is equivalently written as

$$Y = MZ + \varepsilon = MBB^{-1}Z + \varepsilon = \widetilde{M}\widetilde{Z} + \varepsilon, \tag{2.14}$$

where $\widetilde{M} = MB$ and $\widetilde{Z} = B^{-1}Z$. Based on the observation matrix Y it is impossible to distinguish loadings $\widetilde{M}$ from M and factors $\widetilde{Z}$ from Z.

In order to make factor models more feasible some statistical properties are often assumed. One of the most popular models are *orthogonal factor model*, where the factors Z_t are assumed to have zero mean and identity covariance matrix. Additionally, they are uncorrelated to specific factors ε_t, which have also zero mean and covariance matrix Σ_ε. Such a setting allows to find easier interpretation of the factors. In order to increase interpretation power of the model one may also search for such a matrix B that the variance of a squared factors or loadings is maximal. Another possibility is to project them on some known fixed spaces. If $\breve{Z}$ ($\breve{M}$) is specified factors (loading) matrix one may search for a matrix B such that $\| B^{-1}Z - \breve{Z} \|$ ($\| BM - \breve{M} \|$) is minimal, where $\| \cdot \|$ stands for the Euclidean norm.

2.4.2 Principal Component Analysis

There are several methods for estimating factor models. We focus on the method based on a principal component analysis (PCA). For other possible approaches we refer to Johnson and Wichern (1998).

Although the PCA can be treated as a statistical technique of its own rights, here we present it mainly as factorial method. It serves as a tool for

identification factors in orthogonal factor framework and for such a purpose is used in the following discussion.

Let Σ be a covariance matrix of Y_t and let it has a spectral decomposition $\Sigma = \Gamma\Lambda\Gamma^\top$, where $\Gamma = (\gamma_1, \ldots, \gamma_J)$ is $(J \times J)$ orthogonal matrix of eigenvectors and Λ is a diagonal matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_J$ on a diagonal. The principal components are defined as

$$PC_t = \Gamma^\top Y_t,$$

where PC_t is $(J \times 1)$ vector. Note that PC_t is just a linear combination of Y_t coordinates. The covariance matrix of PC_t is Λ, which means that its coordinates are uncorrelated with the variance given by $\lambda_1, \ldots, \lambda_J$.

Obviously $Y_t = \Gamma PC_t$. This shows that in the PCA the vector Y_t is simply transformed to the new basis. Let for the moment $L = J$, then (2.12) is obtained with zero variance of specific factors. However, it is not the interesting case for factor analysis since the aim is dimension reduction. In order to achieve it for $L < J$, set M as first L columns of Γ i.e. $M = (\gamma_1, \ldots, \gamma_L)$ and Z_t as first L rows of PC_t i.e. $Z_t = (PC_t^{(1)}, \ldots, PC_t^{(L)})$. Note that this estimation procedure accomplish the orthogonal factor model assumptions.

The natural question how to choose the number of factors can be answered by explained variance of L factors. One calculates $EV(L) = \frac{\sum_{l=1}^{L} \lambda_l}{tr(\Lambda)}$ and sets L such that $EV(L)$ is sufficiently large. Another possibility is to check when the increments of $EV(L)$ are small. The model choice criteria based on this idea are introduced in Kaiser (1960) and Cattell (1966) among others. Similar treatment may be performed for the residual variation $RV(L) = \frac{1}{TJ} \sum_t \parallel Y_t - MZ_t \parallel^2$.

2.5 Time Series

In our previous discussion we did not look at how the data are collected. We simply took information given in the form it arrived. However, in many situations the data comes in a time sequence and the statistical modelling tries to process the information using a time arrival structure. Having model on the system's dynamics one may try to deduce some information on future states of nature.

The convenient modelling tool for time dependent data are stochastic processes. The stochastic process is a random function $Y : \mathbb{T} \times \Omega \to R^J$, where $\mathbb{T}$ is a ordered set. Here we discuss mainly discrete processes, which means that $\mathbb{T}$ is countable. Mostly positive integers are considered.

Let $Y_t \in \mathbb{R}^J$ be observed for $t = 1, \ldots, T$. Here the index t reflects dynamics of the observations. In practice it means that the random experiment Y_t occurred later in time than Y_{t-1}, Y_{t-1} came after Y_{t-2} and so forth. Typically the index t is associated with regular time measurements (e.g. daily prices), but it is not always the case. The real time increments between different measurements do not always coincide. An standard example from financial markets is exclusion of the weekends or holidays, when there is no public trade so no price is recorded. The indexing reflects then next available record regardless different real time spanning.

The aim of this section is to present some basic results on time series analysis, which are later used for the further discussion. We focus on vector autoregressive processes (VAR) and present basic related concepts, properties and some extensions. For more detailed discussion we refer to Lütkepohl (1993, 2004).

2.5.1 Vector Autoregressive Processes

The main concept of the time series analysis focuses on dependence of the new record on the recent observation. If these dependence is neglected one gains nothing by treating data in the time series context. The dependence structure is most commonly described by a linear function, which leads to the vector autoregressive models. The VAR(p) model can be written as

$$Y_t = \mathcal{A}_1 Y_{t-1} + \ldots + \mathcal{A}_p Y_{t-p} + V + U_t, \tag{2.15}$$

where p is the number of lags, $\mathcal{A}_i$ is a $(J \times J)$ matrix of coefficients and V is a $(J \times 1)$ vector of constants. Finally, U_t is an i.i.d. zero mean vector innovation process with covariance matrix Σ_U. The univariate case $(J = 1)$ is simply referred as AR processes. Note that if for each $i = 1, \ldots, p$ matrix $\mathcal{A}_i$ is diagonal then one obtains collection of AR processes, which can be in fact modelled separately.

The process is called stable if

$$\det(I_J - \mathcal{A}_1 y - \ldots - \mathcal{A}_p y^p) \neq 0 \quad \text{for} \quad |y| < 1.$$

The process is called stationary if its first and second moments are time invariant. In other words it means that $EY_t = \mu$ and

$$E(Y_t - \mu)(Y_{t-h} - \mu)^\top = \Gamma_h$$

are time invariant for each $h = 0, 1, \ldots$ and the matrix Γ_h is referred as autocovariance matrix of the process Y_t. For $h = 0$ one obtains traditional

symmetric covariance matrix. Sometimes it is convenient to consider $\Gamma_{-h} = \Gamma_h^\top = E(Y_{t-h} - \mu)(Y_t - \mu)^\top$.

The stationary process (2.15) can be equivalently written in a mean adjusted form

$$Y_t - \mu = \mathcal{A}_1(Y_{t-1} - \mu) + \ldots + \mathcal{A}_p(Y_{t-p} - \mu) + U_t.$$

Assume for simplicity that Y_t is a stationary zero mean process. An important characteristic of a process is its covariance structure, which we understand as collection of matrices Γ_h. Note that

$$\Gamma_h = \mathcal{A}_1\Gamma_{h+1}^\top + \ldots + \mathcal{A}_p\Gamma_{h+p}^\top + \Sigma_U,$$

which suggest the existence of some correspondence between the covariance structure and the parameter of the process. For the illustration purpose consider the VAR(1) process. It is easy to show that

$$\Gamma_0 = \mathcal{A}_1\Gamma_0\mathcal{A}_1^\top + \Sigma_U,$$

which can be written equivalently with the stack form as

$$\gamma_0 = (I_{J^2} - \mathcal{A}_1 \otimes \mathcal{A}_1)^{-1}\sigma_U,$$

where γ_0 and σ_U denote a column stack form of Γ_0 and Σ_U respectively.

2.5.2 Estimation of VAR Models

In the previous section we discussed some properties of VAR processes. This part is devoted to the estimation of the VAR parameters. First we note that the mean and the autocovariances can be consistently estimated with $\widehat{\mu} = \bar{Y}_t$ and

$$\widehat{\Gamma}_h = \frac{1}{T}\sum_{t=h+1}^{T}(Y_t - \bar{Y}_t)(Y_{t-h} - \bar{Y}_t)^\top.$$

We recall that Y_t is treated as a zero mean vector, which can in practice be achieved by substraction of the sample mean.

In order to obtain the estimators for the VAR(p) process parameters define $Y = (Y_{p+1}, \ldots, Y_T)$, $U = (U_{p+1}, \ldots, U_T)$ which are $(J \times T - p)$ matrices and $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_p)$, which is a $(J \times pJ)$ matrix. Additionally let $\mathcal{Y}_t = (Y_t^\top, \ldots, Y_{t-p+1}^\top)^\top$ and $\mathcal{Y} = (\mathcal{Y}_{p+1}, \ldots, \mathcal{Y}_{T-1})$, which is $(pJ \times T - p)$ matrix. Then in the matrix notation the VAR model can be written

$$Y = \mathcal{A}\mathcal{Y} + U.$$

The least square estimator is now given as

$$\widehat{\mathcal{A}} = Y\mathcal{Y}^\top(\mathcal{Y}\mathcal{Y}^\top)^{-1}. \tag{2.16}$$

Let $\widehat{\theta}$ denotes the stack form of $\widehat{\mathcal{A}}$ and θ the corresponding stack form of $\mathcal{A}$. Then it can be shown that

$$\sqrt{T}(\widehat{\theta} - \theta) = \mathcal{O}_P(1). \tag{2.17}$$

The estimator (2.16) can be represented in terms of the covariance structure. Note that $Y\mathcal{Y}^\top/T$ is a $(J \times Jp)$ matrix of a form $\widehat{\Gamma}_Y = \left(\widehat{\Gamma}_1, \ldots, \widehat{\Gamma}_p\right)$ and $\widehat{\mathcal{Y}}\widehat{\mathcal{Y}}^\top/T$ is given as

$$\widehat{\Gamma}_{\mathcal{Y}} = \begin{pmatrix} \widehat{\Gamma}_0 & \widehat{\Gamma}_1 & \ldots & \widehat{\Gamma}_p \\ \widehat{\Gamma}_1^\top & \widehat{\Gamma}_0 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\Gamma}_p^\top & \widehat{\Gamma}_{p-1}^\top & \ldots & \widehat{\Gamma}_0 \end{pmatrix}. \tag{2.18}$$

Thus the estimator of $\mathcal{A}$ can be rewritten to

$$\widehat{\mathcal{A}} = \widehat{\Gamma}_Y \widehat{\Gamma}_{\mathcal{Y}}^{-1}. \tag{2.19}$$

This form of the estimator is often referred as Yule-Walker estimator. Here we want to emphasize that the knowledge of covariance structure yields automatically the information about the parameters.

2.5.3 Autocorrelation Tests

The previous section is devoted to the estimation of the VAR model. It presents the methodology for estimating the parameters when the structure of the model is assumed. However it may be worth to apply some statistical diagnostic tools in order to validate how adequate is the fitted model.

One of the possible checking procedure is based on testing for autocorrelation of residuals. In Section 2.5.1 we assume U_t to be a white noise process and if it is not the case the model may be misspecified and lead to wrong inference. Therefore, one applies the autocorrelation tests to empirical residuals defined as

$$\widehat{U}_t = Y_t - \widehat{\mathcal{A}}_1 Y_{t-1} - \ldots - \widehat{\mathcal{A}}_p Y_{t-p}$$

for $t = p+1, \ldots, T$. If U_t is in fact the white noise then the autocovariance matrices of the residuals denoted here as Σ_{U_h} for $h = 1, 2, \ldots$, are zero. This fact may be tested on the empirical matrices $\widehat{\Sigma}_{U_h}$.

Here we present two test for residual autocorrelation, which are popular in applied work, namely portmanteau test and Lagrange multiplier test (LM). The first one tests the hypothesis of zero autocovariances

$$\begin{aligned} H_0: &\quad \Sigma_{U_i} = 0 \quad \text{for all} \quad i = 1, \ldots, h \\ H_1: &\quad \Sigma_{U_i} \neq 0 \quad \text{for some} \quad i = 1, \ldots, h. \end{aligned}$$

The test statistic has form

$$Q_p = T \sum_{i=1}^{h} \operatorname{tr}(\widehat{\Sigma}_{U_i}^{\top} \widehat{\Sigma}_U^{-1} \widehat{\Sigma}_{U_i} \widehat{\Sigma}_U^{-1}),$$

where all estimators are obtained with standard estimation techniques. The statistics Q_p has approximately asymptotic χ^2 distribution, see Ahn (1988). The approximation holds for h going to infinity together with the sample size. One considers also adjusted portmanteau statistics

$$\widetilde{Q}_p = T^2 \sum_{i=1}^{h} \frac{1}{T-i} \operatorname{tr}(\widehat{\Sigma}_{U_i}^{\top} \widehat{\Sigma}_U^{-1} \widehat{\Sigma}_{U_i} \widehat{\Sigma}_U^{-1}),$$

with superior small sample size properties, which has same asymptotic properties as Q_p.

The idea of the LM test is to assume that the error terms follow the VAR(h) model itself

$$U_t = \mathcal{B}_1 U_{t-1} + \ldots + \mathcal{B}_h U_{t-h} + \widetilde{U}_t,$$

and test the hypothesis

$$\begin{aligned} H_0: &\quad \mathcal{B}_i = 0 \quad \text{for all} \quad i = 1, \ldots, h \\ H_1: &\quad \mathcal{B}_i \neq 0 \quad \text{for some} \quad i = 1, \ldots, h. \end{aligned}$$

Test statistic Q_{LM} has a asymptotic approximate χ^2 distribution and has relatively complex form. Therefore, we do not present it here but refer to Lütkepohl (2004). The alternative procedure is to test whether only one of $\mathcal{B}_i = 0$. The test statistic remains to be χ^2 distributed. Note that now procedure of testing of zero autocorrelation up to order h has a different size due to the sequential testing problems.

The two presented tests have embedded h parameter, which have to be set. In order to study small sample properties one has to apply Monte Carlo simulations. Details can be found in Hosking (1980, 1981) or Brüggemann,

Lütkepohl, and Saikkonen (2006). Here we state that LM has satisfactory properties only if autocorrelation of small order is tested. For the portmanteau test one has to test higher order residual autocorrelation to get correct size in a small sample.

To close this section we need to mention that the presented tests check only autocorrelation of the returns. However, the white noise may be violated differently, e.g. conditional heteroscedastic structure. This effects are not considered in this work.

2.5.4 Forecasting

Forecasting is one of the main aims of the time series analysis. Based on the information up to time T one would like to predict the state of the system h steps ahead i.e. in the future time $T + h$. In fact the predictor is given as an outcome off some function ϕ, which operates on $Y_1, \ldots, Y_T$ only.

In order to construct the optimal function ϕ one has to define the loss of the forecast and minimize this loss among all possible functions ϕ. The most common approach is to assume mean squared loss. Then it can be proven that optimal forecast h steps ahead denoted by Y^*_{T+h} is given by conditional expected value

$$y^*_{T+h} = E(Y_{T+h}|Y_T = y_T, \ldots, Y_1 = y_1).$$

Here for the moment we assume that Y_t is a random variable and y_t its observed realization.

In the case of VAR(p) models the predictors satisfy a recursive relation

$$\begin{aligned} Y^*_{T+1} &= \mathcal{A}_1 Y_T + \ldots + \mathcal{A}_p Y_{T-p+1} \\ Y^*_{T+2} &= \mathcal{A}_1 Y^*_{T+1} + \ldots + \mathcal{A}_p Y_{T-p+2} \\ &\vdots \end{aligned}$$

Note that it utilizes the fact that U_t is a zero mean white noise.

The forecast discussed above is a point forecast, since the procedure yields exactly one element for Y^*_{T+h}. The quality of the prediction can be evaluated by analyzing the realized forecast error defined by some loss function. Apart from point forecasts one may also study interval or distributional forecasts. Then, the conditional expected value of the form

$$Y^*_{T+h} = E(Y_{T+h}|Y_T, \ldots, Y_1).$$

yields a random variable. Evaluation of the interval forecast is however more demanding. In the VAR(p) example some distribution of U_t has to be additionally assumed in order to obtain the forecast.

2.5.5 Dynamic Factor Models

In case when the dimension J of the time series Y_t is relatively large it may be advisable to apply some dimension reduction techniques. One possible approach utilize the factor analysis discussed in Section 2.4. Y_t can be then rewritten to

$$Y_t = MZ_t + \varepsilon_t, \tag{2.20}$$

which is exactly (2.12). This representation assumes existence of comovements among all component of Y_t, which are driven by unobservable factors Z_t.

In (2.20) Z_t itself can be treated as time series. In particular VAR model for factors can be assumed, which yields

$$\begin{aligned} Y_t &= MZ_t + \varepsilon \\ Z_t &= \mathcal{A}_1 Z_{t-1} + \ldots + \mathcal{A}_p Z_{t-p} + U_t. \end{aligned} \tag{2.21}$$

As dynamics of the factors is incorporated we may call it dynamic factor model. These kind of models are widely used in empirical macroeconomics, see Stock and Watson (2005) or Breitung and Eickmeier (2005). In fact the usual exposition of dynamic factor models has more complex structure. In compare to (2.21) lags in factors and series itself are included i.e.

$$\begin{aligned} Y_t &= M_0 Z_t + \ldots + M_q Z_{t-q} + \mathcal{B}_1 Y_{t-1} + \ldots + \mathcal{B}_r Y_{t-r} + \varepsilon_t \\ Z_t &= \mathcal{A}_1 Z_{t-1} + \ldots + \mathcal{A}_p Z_{t-p} + U_t. \end{aligned}$$

Often the moving average representation is used, see Chapter 1, Forni et al. (2000), Forni and Lippi (2001). Details concerning estimation, identification and application to the macroeconomic analysis can be found in Stock and Watson (2005).

2.6 Panel Data Analysis

Section 2.5 discusses the times series analysis, where one observes the evolution of measurements over time. However, for each observed object some

additional characteristics could be given. The data observed across sections and over time are called *panel data* or *longitudinal data*. The modelling of these structures extends regression models by additional time dynamics. On the other hand, it refines as well time series analysis by including supplementary information.

Let a pair (X_{jt}, Y_{jt}) be observed for $t = 1, \dots, T$ and $j = 1, \dots, J$, where $X_{jt} \in \mathbb{R}^d$ and $Y_{jt} \in \mathbb{R}$. Here j indicates the index of the object and t its time evolution. Note that this setting can be either split to J time series of $(d+1)$-dimensions or T separate regression problems with J observations. However, analyzing simultaneously in space and time dimensions allows to recover effects that may be otherwise hidden for only cross sectional or time series data.

In the previous notation we implicitly assume some regular structure. However, not every object has to be observed on the same time domain. In practice it is common that for j-th observations only some values on the subset $\mathcal{T} \subset \{1, \dots, T\}$ are available. This case we call *unbalanced panel* in contrast to *balanced panel* of regular design.

Below we shortly discuss the methods of longitudinal data analysis. For more details we refer to Baltagi (2005) or Hsiao (2003). The methods presented here are parametric, for nonparametric methods we refer to Wu and Zhang (2006).

2.6.1 Fixed Effects Model

One of the most common modelling approach of the panel data is *fixed effect model*. The dependent variable is explained by the time dependent regressors and unobservable object specifics, which do not vary in time. The regression function is linear and has the following form

$$Y_{jt} = V_j + \sum_{i=1}^{d} Z^{(i)} X_{jt}^{(i)} + \varepsilon_{jt}, \tag{2.22}$$

where V_j denotes the objects specific, $Z = (Z^{(1)}, \dots, Z^{(d)})^\top$ regression coefficients and ε_{jt} zero mean with finite second moment i.i.d. errors. As before $X_{jt}^{(i)}$ is the ith coordinate of X_{jt} and is independent of $\varepsilon_{j't'}$. Note that V_j does not depend on time and therefore reflects the unobservable feature of the object j. The common structure for all objects is represented by Z loaded with observable characteristics X_{jt}. In fixed effects model the common structure does not vary in time and the whole dynamic behavior is explained only by variation in observable features.

The model (2.22) can be conveniently written in a matrix notation

$$Y = D_V V + XZ + \varepsilon, \tag{2.23}$$

where $Y = (Y_{11}, Y_{12}, \ldots, Y_{1T}, \ldots, Y_{JT})^\top$ is a $(JT \times 1)$ vector of dependent variables, $X = (X_{11}, X_{12}, \ldots, X_{JT})^\top$ is $(JT \times d)$ matrix and $V = (V_1, \ldots, V_J)^\top$ is $(J \times 1)$ vector. D_V is a $(JT \times J)$ matrix containing only ones and zeros which can be written $D_V = I_J \otimes 1_T$, where I_J is an identity matrix of dimension J and 1_T is a $(T \times 1)$ vector of ones.

The unknown parameter to estimate in (2.23) is $\theta = (V, Z)$. Note that (2.23) might be rewritten to

$$Y = (D_V, X)(V^\top Z^\top)^\top + \varepsilon.$$

This representation suggest using the least square estimator discussed in Section 2.3.1. For simplicity we assume that the matrix (D_V, X) is of a full rank. However, the inversion of (D_V, X) may cause numerical difficulties when $J + T$ is too large.

The numerical problems lead to more feasible sequential approach. The first step is to transform the data in order to remove the groups' means. Then estimating Z is reduced to d-dimensional least square problem. In the final step the estimate of V is calculated.

For substraction of V define the symmetric idempotent matrix $W_V = I_{JT} - D_V(D_V^\top D_V)^{-1} D_V^\top$. Note that multiplying (2.23) by W_V leads to

$$W_V Y = W_V XZ + W_V \varepsilon, \tag{2.24}$$

since $W_V D_V V = 0$. The least square estimator for Z is

$$\widehat{Z} = (X^\top W_V X)^{-1} X^\top W_V Y. \tag{2.25}$$

Consequently the estimator for V is

$$\widehat{V} = (D_V^\top D_V)^{-1} D_V^\top (Y - X\widehat{Z}). \tag{2.26}$$

The estimator $\widehat{Z}$ is consistent. Moreover for $T \to +\infty$ and J fixed, $\widehat{V}$ is also a consistent estimator. However, it is not the case when $J \to +\infty$ since the dimension of V converges to infinity as well.

The estimation idea for unbalanced panel is basically same. The matrix D_V has to be carefully redefined since the number of ones is not necessarily constant in each column.

2.6.2 Random Effects Model

In fixed effects model (2.22) a parameter V_j is associated to each particular individual. This however may cause the overparameterization of the model. A parallel approach is to treat the unobserved unit's specifics as a random variable. It is also feasible to assume a randomness of the specifics for the inference on individuals, which are rather drawn from a large population. A classical example is a household study, where one treat the observations as a representative sample and is not interested of the results fixed to this particular sample.

In the random effects model

$$V_j = \widetilde{V} + \widetilde{\varepsilon}_j,$$

where $\widetilde{\varepsilon}_j$ are i.i.d zero mean with σ_V^2 variance. $\widetilde{V}$ reflects now the specific common variable for all units. The model (2.22) can be simply rewritten to

$$Y_{jt} = \widetilde{V} + \sum_{i=1}^{d} Z^{(i)} X_{jt}^{(i)} + (\widetilde{\varepsilon}_j + \varepsilon_{jt}) = \sum_{i=0}^{d} Z^{(i)} X_{jt}^{(i)} + \widetilde{\widetilde{\varepsilon}}_{jt}, \tag{2.27}$$

where for convenience of the notation we set $Z^{(0)} = \widetilde{V}$, $X_{jt}^{(0)} = 1$ and $\widetilde{\widetilde{\varepsilon}}_{jt} = \widetilde{\varepsilon}_j + \varepsilon_{jt}$. We assume that $\widetilde{\varepsilon}_j$ and ε_{jt} are pairwise uncorrelated and ε_{jt} are i.i.d. zero mean with variance σ_ε^2. Therefore, the covariance matrix of $\widetilde{\widetilde{\varepsilon}}_{jt}$ is not a diagonal matrix but has the following form

$$\Sigma_\varepsilon = \sigma_V^2 (I_J \otimes 1_{T\times T}) + \sigma_\varepsilon^2 I_{TJ}, \tag{2.28}$$

where $1_{T\times T}$ is $(T \times T)$ matrix of ones.

In the matrix notation the model has a form

$$Y = XZ + \widetilde{\widetilde{\varepsilon}}, \tag{2.29}$$

where Y, X, Z and $\widetilde{\widetilde{\varepsilon}}$ are matrices constructed from elements of (2.27). The parameter Z can be estimated by generalized least squares

$$\widehat{Z} = (X^\top \Sigma_\varepsilon^{-1} X)^{-1} X^\top \Sigma_\varepsilon^{-1} Y. \tag{2.30}$$

In order to obtain Σ_ε^{-1} some calculation trick is commonly advised. Note that (2.28) can be rewritten to

$$\Sigma_\varepsilon = (T\sigma_V^2 + \sigma_\varepsilon^2) \left(I_T \otimes \frac{1}{T} 1_{J\times J} \right) + \sigma_\varepsilon^2 \left\{ I_{TJ} - \left(I_J \otimes \frac{1}{T} 1_{T\times T} \right) \right\}. \tag{2.31}$$

One can easily check that $W_{\widetilde{V}} = I_J \otimes \frac{1}{T} 1_{T\times T}$ and $W_V = I_{TJ} - (I_J \otimes \frac{1}{T} 1_{T\times T})$ are symmetric and idempotent. Moreover $W_V + W_{\widetilde{V}} = I_{TJ}$. This yields

$$\Sigma_\varepsilon^{-1} = \frac{1}{T\sigma_V^2 + \sigma_\varepsilon^2} W_{\widetilde{V}} + \frac{1}{\sigma_\varepsilon^2} W_V.$$

Now one has to plug-in some estimators $\widehat{\sigma}_V^2$ and $\widehat{\sigma}_\varepsilon^2$. There are several proposals for obtaining these estimators, see Wallace and Hussain (1969), Amemiya (1971), Swamy and Arora (1972), Nerlove (1971). One possibility is to run regression *within-groups* to wipe out V_j terms from (2.27) by averaging like in (2.24). Then, the estimator has a form of

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{J(T-1) - d} \left\{ Y^\top W_V Y - Y^\top W_V X (X^\top W_V X)^{-1} X^\top W_V Y \right\}. \quad (2.32)$$

The next step is the *between* regression averaging across the time. Multiplying (2.29) by $W_{\widetilde{V}}$ one obtains the regression model for the means across the time period, which yields the following equation

$$T\widehat{\sigma}_V^2 + \widehat{\sigma_\varepsilon}^2 = \frac{1}{J - 1 - d} \left\{ Y^\top W_{\widetilde{V}} Y - Y^\top W_{\widetilde{V}} X (X^\top W_{\widetilde{V}} X)^{-1} X^\top W_{\widetilde{V}} Y \right\} \quad (2.33)$$

where $\widehat{\sigma}_\varepsilon^2$ is a given by (2.32).

2.6.3 Time Specific Effects

Sections 2.6.1 and 2.6.2 discuss models where the time innovations comes only from dynamics of exploratory variables X_{jt}. In this section we present the natural extension with a common time component. The model (2.22) is refined to

$$Y_{jt} = V_j + \sum_{i=1}^{d} Z^{(i)} X_{jt}^{(i)} + \Lambda_t + \varepsilon_{jt}, \quad (2.34)$$

where Λ_t is denotes unobservable time effect invariant across individuals. It can be interpreted as external time varying factor, which influence each Y_{jt} in a similar manner and does not need to have a direct impact on X_{jt}. Note that in case of $V_j = 0$ for all j the model (2.34) can be treated like in Sections 2.6.1 and 2.6.2 with interchange of indexes j and t.

In the fixed effects framework (2.23) and be rewritten to

$$Y = D_V V + XZ + D_\Lambda \Lambda + \varepsilon, \quad (2.35)$$

where $\Lambda = (\Lambda_1, \ldots, \Lambda_t)^\top$ is $(T \times 1)$ vector of time effects and $D_\Lambda = 1_J \otimes I_T$. Let $W_\Lambda = I_{TJ} - D_\Lambda (D_\Lambda^\top D_\Lambda)^{-1} D_\Lambda^\top$ be defined similarly to W_V from Section 2.6.1 then $W_\Lambda D_\Lambda \Lambda = 0$. Note that $W_V W_\Lambda = W_\Lambda W_V$.

Multiplying (2.35) by $W_V W_\Lambda$ wipes out both individual and time effects and leads to the estimator

$$\widehat{Z} = (X^\top W_V W_\Lambda X)^{-1} X^\top W_V W_\Lambda Y.$$

By cancelling only time or only individuals effects the corresponding estimators $\widehat{V}$ and $\widehat{\Lambda}$ are easily obtained.

Considering random effects one assumes additionally that Λ_t is i.i.d zero mean with variance σ_Λ^2 and is independent of ε_{jt}, $\widetilde{\varepsilon}_j$ and X_{jt}. $\widetilde{\widetilde{\varepsilon}}_{jt}$ from (2.27) is updated to $\widetilde{\widetilde{\varepsilon}}_{jt} = \widetilde{\varepsilon}_j + \varepsilon_{jt} + \Lambda_t$ and its covariance matrix to

$$\Sigma_\varepsilon = \sigma_V^2 (I_T \otimes 1_{J \times J}) + \sigma_\varepsilon^2 I_{TJ} + \sigma_\Lambda^2 (1_{T \times T} \otimes I_J),$$

which can be transform to a sum with some calculation trick similarly to (2.31). Define the following matrices:

$$\begin{aligned}
W_1 &= \left(I_J - \frac{1}{J} 1_{J \times J}\right) \otimes \left(I_T - \frac{1}{T} 1_{T \times T}\right) \\
W_2 &= \left(I_J - \frac{1}{J} 1_{J \times J}\right) \otimes 1_{T \times T} \\
W_3 &= 1_{J \times J} \otimes \left(I_T - \frac{1}{T} 1_{T \times T}\right) \\
W_4 &= 1_{J \times J} \otimes 1_{T \times T}.
\end{aligned}$$

Note that matrices $W_1, \ldots, W_4$ are symmetric idempotent and pairwise orthogonal. Then

$$\Sigma_\varepsilon = \sigma_\varepsilon^2 W_1 + (T\sigma_V^2 + \sigma_\varepsilon^2) W_2 + (J\sigma_\Lambda^2 + \sigma_\varepsilon^2) W_3 + (T\sigma_V^2 + J\sigma_\Lambda^2 + \sigma_\varepsilon^2) W_4.$$

As a direct consequence the inverse matrix is obtained as

$$\Sigma_\varepsilon^{-1} = \frac{W_1}{\sigma_\varepsilon^2} + \frac{W_2}{T\sigma_V^2 + \sigma_\varepsilon^2} + \frac{W_3}{J\sigma_\Lambda^2 + \sigma_\varepsilon^2} + \frac{W_4}{T\sigma_V^2 + J\sigma_\Lambda^2 + \sigma_\varepsilon^2}.$$

In order to obtain estimators of variances one has to run three regressions by multiplying the model by W_1, W_2, W_3 respectively. The estimators of variances have the similar form to (2.32) or (2.33). The final estimator of Z is also given by (2.30) when the appropriate estimators of variances are plugged in.

2.6.4 Dynamic Panel Models

The main feature of the panel data is its dynamic context. The modelling can catch it by evolution of external variable X_{jt} like in Sections 2.6.1 and 2.6.2 or by introducing additional time dependent variable Λ_t like in Section 2.6.3. Another approach is to allow for lags in Y_{jt}. This modelling strategy is referred as dynamic panel models and share some similarities to classical time series modelling discussed in Section 2.5.

The dynamic panel model has the following form

$$Y_{jt} = \widetilde{V}_j + \sum_{i=1}^{d} Z^{(i)} X_{jt}^{(i)} + \gamma Y_{j,t-1} + \varepsilon_{jt}.$$

Treating $Y_{j,t-1}$ as a next observable exploratory variable and γ as a additional parameter to estimate the methodology from Sections 2.6.1 and 2.6.2 can be adopted. One has to carefully adjust the matrices in (2.23) and (2.29) because of the lag in Y_{jt}.

Although the estimation procedure can be adjusted for the lags the statistical properties of the estimators are different. $(\widehat{Z}^\top, \widehat{\gamma})^\top$ is no longer consistent, because $Y_{j,t-1}$ is correlated with $\varepsilon_{j,t-1}$. To avoid this problems one may use instrumental variable e.g. build model for differences.

2.7 Functional Data Analysis

In Section 2.2 we define a random variable X as measurable function from some probability space $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}^d, \mathcal{B}^d)$. Being more general $\mathbb{R}^d$ together with standard scalar product is a classical example of Hilbert space, for more details on Hilbert spaces see Rudin (1991). Another popular example of Hilbert space is the space of Lebesgue square integrable functions - $L_2(\Xi)$ on some subset $\Xi \subseteq \mathbb{R}^d$ with the scalar product $< f_1, f_2 >= \int_\Xi f_1(u) f_2(u) du$. As natural extension to a random variable one may consider measurable functions from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{H}, \mathcal{B}_\mathbb{H})$, where $\mathbb{H}$ is a Hilbert space, possibly space of functions, and $\mathcal{B}_\mathbb{H}$ is an appropriate σ-field. In particular when $\mathbb{H}$ is a function space this type of modelling we call *functional data analysis*.

This section presents only few results on functional data analysis. We will use them in the next sections. More comprehensive view on this topic can be found in the monograph of Ramsay and Silverman (1997). The discussion on financial applications of the functional data analysis can be found in Benko (2006).

The fundamental question of the functional data analysis is why should one treat the observations in the functional context. In fact if the domain of

the functions is a compact subset of $\mathbb{R}^d$ it would be never possible to observe all values of a particular function but only its finite sub-sample. However, exploring some features that are meaningful only in the functional context, e.g. smoothness, may yield additional information. Suppose we observe unbalanced panel structure but the individuals are not sampled exactly at the same time points. For example, consider a hypothetical children growth data, where the height measurement are done similar but not exactly the same age. It seems to be reasonable to assume the height being a smooth function of the age and the growth acceleration could be conveniently studied by taking a derivative of this function.

2.7.1 Principal Components for Functional Data

Here we present some basic concept of functional data analysis namely functional principal components analysis (fPCA). It is a natural extension for functional data of the PCA discussed in Section 2.4.2. Similarly to multivariate case we are mainly interested in the fPCA as factorial method.

Let $Y_1(x), \ldots, Y_T(x) \in L_2(\Xi)$ be random functions with $\Xi \subseteq \mathbb{R}^d$. In order to simplify notation we assume that $EY_t = 0$. The expectation is understood here in functional sense. One may also extend the notation of covariance matrix to covariance function on cartesian product $\Xi \times \Xi$

$$\Sigma(x_1, x_2) = E(Y_t(x_1)Y_t(x_2)).$$

We assume here the necessary regularities of the random functions Y_t, i.e. existence of cross-moments for each pair $x_j, x_{j'} \in \Xi$. Define the covariance operator $\mathbf{\Gamma}$ operating on function $V \in L_2(\Xi)$

$$(\mathbf{\Gamma}V)(x_1) = \int_\Xi \Sigma(x_1, x_2)V(x_2)dx_2.$$

From Fredholm theorem the operator $\mathbf{\Gamma}$ has at most countable eigenvectors $\gamma_1(x), \gamma_2(x), \ldots$. Moreover eigenvalues have to converge to zero, so one is able to form a non-increasing sequence of eigenvalues $\lambda_1 \geq \lambda_2 \ldots$. Then $\Sigma(x_1, x_2) = \sum_i \lambda_i \gamma_i(x_1)\gamma_i(x_2)$. It corresponds to spectral decomposition of covariance matrix in a multivariate framework.

The notion of functional principal components can be regarded in a similar manner. $PC_t = (PC_t^{(1)}, PC_t^{(2)}, \ldots)$ is an infinite series, where

$$PC_t^{(i)} = \int_\Xi \gamma_i(x)Y_t(x)dx.$$

Note that the scalar product given by the integral replace the summation in multivariate case.

2.7.2 Karhunen-Loève Expansion

The Karhunen-Loève expansion, known also as Karhunen-Loève decomposition, gives a convenient series representation of a random function. It shares the same principals as Fourier series expansion for deterministic functions. In the deterministic case a function can represented by a series of scalars loaded with some predefined basis functions, see Section 2.1. Considering the random functions the representation is given as a series of scalar random variables loaded with deterministic basis functions. More precisely the random function Y_t can be represented as

$$Y_t(x) = \sum_{l=1}^{+\infty} Z_{tl} m_l(x),$$

where Z_{tl} are random variables and $m_l(x) \in L_2(\Xi)$ are non-random basis functions.

The Karhunen-Loève expansion follows the idea of the PCA and the fPCA. The functions m_l are defined as eigenfunctions of covariance operator Γ i.e. $m_l = \gamma_l$, and Z_{tl} as a scalar product of m_l and Y_t i.e. $Z_{tl} = PC_t^{(l)}$. Because of its nice properties this representation is of great importance. From Section 2.7.1 one may deduce that m_l are pairwise orthogonal functions and Z_{tl} are pairwise uncorrelated. For our analysis the most significant issue is the *best empirical basis property*. It says that for any fixed L the finite approximation given by the Karhunen-Loève expansion is optimal in sense of mean square error, i.e.

$$(PC_t, \gamma) = \arg\min_{(Z_t, m)} E \int_{\Xi} \left\{ Y_t(x) - \sum_{l=1}^{L} Z_{lt} m_l(x) \right\}^2 dx,$$

where $PC_t = (PC_t^{(1)}, \ldots, PC_t^{(L)})^\top$, $\gamma = (\gamma_1, \ldots, \gamma_L)^\top$ is a tuple of L basis functions, and similarly $Z_t = (Z_{t1}, \ldots, Z_{tL})^\top$ $\gamma = (m_1, \ldots, m_L)^\top$. The best empirical basis property suggest immediately the factorial usage of the fPCA.

2.7.3 Estimation

Up to now we discuss the theoretical framework of operating on the random functions. However, in practice the functions operating on continuous domain can be never fully observed. Instead one sees some finite number of function's values together with some error. Therefore to utilize the presented methodology one has to use certain approximations, discussed already in Section 2.1. As a consequence the estimation in functional framework often boils down to the estimation in multivariate framework.

Consider the random functions $Y_t(x)$ and let (X_{tj}, Y_{tj}) be observed for $t = 1, \ldots, T$ and $j = 1, \ldots, J_t$. Applying interpolation or smoothing techniques one may obtain the estimates of $\widehat{Y}_t(x)$ for each particular t. The quality of the estimator depends on the smoothing level, design of the points X_{tj} and the number of points J_t. In fact as it was already discussed before there are two main ideas of the finite representation of a function, namely discrete representation on a grid and series expansion. Below we present how to proceed in these two schemes in case of the fPCA.

Let $\mathcal{G} = \{x_1, \ldots x_N\}$ be a sufficiently dense equally spaced grid that span Ξ. Then $\widehat{Y}_t(\mathcal{G})$ can be easily represented in a stack form

$$\widehat{Y}_t(\mathcal{G}) = (\widehat{Y}_t(x_1), \ldots, \widehat{Y}_t(x_N))^\top. \tag{2.36}$$

Define similarly to Section 2.4 matrix $Y = (\widehat{Y}_1(\mathcal{G}), \ldots, \widehat{Y}_T(\mathcal{G}))$. Here Y is $(N \times T)$ matrix. Applying the PCA to Y yields estimates $\widehat{\Gamma}$ and $\widehat{PC}_t$, where $\widehat{\Gamma}$ is a $(N \times N)$ matrix and $\widehat{PC}_t$ is $(N \times 1)$ vector. In order to return to the functional context the columns of $\widehat{\Gamma}$ can be regarded as discrete representation of the functions γ_i on $\mathcal{G}$. Hence the estimates $\widehat{\gamma}_i$ are obtained. The values for any given x have to be computed using some interpolation/extrapolation methods.

A parallel approach to discrete representation uses basis expansion. Suppose that each function is approximated by K basis functions

$$Y_t(x) = \sum_{k=1}^{K} c_{tk}\psi_k(x). \tag{2.37}$$

In more compact form (2.37) can be written

$$\mathcal{Y} = C\psi,$$

where $\mathcal{Y} = (Y_1, \ldots, Y_T)$ and $\psi = (\psi_1, \ldots, \psi_T)$ are tuples of appropriate functions. The covariance function is then estimated by

$$\widehat{\Sigma}(x_1, x_2) = \frac{1}{T}\mathcal{Y}(x_1)^\top \mathcal{Y}(x_2) = \frac{1}{T}\psi(x_1)^\top C^\top C\psi(x_2).$$

Define $(K \times K)$ symmetric matrix W by its elements

$$W = \left(\int_\Xi \psi_{k_1}(x)\psi_{k_2}(x)dx\right)_{k_1 k_2}.$$

Series expansion can be also applied to the eigenfunctions

$$\gamma_i(x) = \sum_{k=1}^{K} a_{ik}\psi_k(x) = \psi^\top(x)a_i,$$

where $a_i = (a_{i1}, \ldots, a_{iK})^\top$. This yields

$$(\Gamma\gamma_i)(x_1) = \int_\Xi \Sigma(x_1, x_2)\gamma_i(x_2)dx_2 = \frac{1}{T}\psi(x_1)^\top C^\top CWa_i$$
$$= \lambda_i\gamma_i(x) = \lambda_i\psi(x)a_i. \tag{2.38}$$

Since it hold for x (2.38) simplifies to the matrix equation

$$\frac{1}{T}C^\top CWa_i = \lambda_i a_i. \tag{2.39}$$

In order to achieve orthonormality of γ_i one applies restriction $a_{i_1}Wa_{i_2} = \delta_{i_1 i_2}$. Therefore (2.39) is transformed by $a_i = W^{-\frac{1}{2}}\widetilde{a}_i$ to the eigenvalue problem

$$\frac{1}{T}W^{\frac{1}{2}}C^\top CW^{\frac{1}{2}}\widetilde{a}_i = \lambda_i\widetilde{a}_i.$$

Note that matrix C has to be estimated from the data and to calculate W a numerical or analytical methods should be applied conditionally on the form of ψ. In particular when ψ_i form the orthonormal basis $W = I_K$.

2.8 Dynamic Regression Problems

In this section we discuss the data of a form (X_{tj}, Y_{tj}) for $j = 1, \ldots, J_t$ and $t = 1, \ldots, T$. We assume $X_{tj} \in \mathbb{R}^d$ and $Y_{tj} \in \mathbb{R}$. The index t reflects the time evolution while j records the observation on the particular time point t. The number of the observations J_t and design of points X_{tj} may depend on time. We call this structure *dynamic regression problem*, which can be modelled by

$$E(Y_t|X_t) = F_t(X_t),$$

where X_t, Y_t are random variables describing the observations at time point t and F_t is some regression function also depending on time.

Similar data structure is given in Sections 2.6 and 2.7. Before we move our discussion to the methods for dynamic regression we shortly recall the differences.

Panel data, in particular unbalanced panel, may be seen as a special case of dynamic regression. One observes the dynamic evolution of some records and the models for longitudinal data from Section 2.6 incorporate these individual effects. In (2.22) or (2.27) they are represented by the term V_j. In dynamic regression problems the observed data points does not need

to posses any external linkage. It means in practice that there is no direct connection between points X_{tj} and $X_{t'j'}$ for $t \neq t'$ or this connection is not used in modelling. It results in setting $V_j = 0$.

In functional data analysis one assumes that some functions $Y_t(x)$ are given. The index t does not have to reflect the time evolution of the functions but simply indicates different records. As an example consider some weather measurements like temperature across one year for different weather stations. They form some periodic functions of x denoting time measurement and index t reflects different stations. However, assuming that t describes time dynamics and x is some space variable, e.g. term structure of interest rate where x is time to maturity, one may apply the methodology of functional data analysis for solving dynamic regression problems. This approach is briefly discussed in Section 2.8.2.

In time varying regression one faces a problem of simultaneous analysis in space and time dimensions. Note that neglecting the time structure would lead to the regression based on the pooled data and the methods from Section 2.3 could be applied. Such a simplification, however, could cause a loss of important information. Therefore, one would rather perform the space regression over time and certain modelling assumptions would build a dynamic linkage. For example, in parametric regression some parameters could be time dependent.

In this Section we discuss mainly models of additive form

$$E(Y_t|X_t) = \sum_{l=0}^{L} Z_{tl} m_l(X_t). \tag{2.40}$$

This representation is relatively parsimonious and allows convenient separation between space and time effects. Time changes are caught by scalars Z_{tl}, which can be afterwards analyzed through times series methods like in Section 2.5. The space dependence is achieved by functions m_l, which operate on exploratory variables X_t and do not depend on time. The models assume the linear link between time and space components. Note that (2.40) can be regarded as a refinement of factor model (2.11) with functional loadings. Z_{tl} are common for all observations in moment t and functions $m_l(X_t)$ determine their influence on Y_t.

In the following discussion we show different possible implementations of (2.40). Section 2.8.1 presents models with parametric functions m_l. In Section 2.8.2 we recall functional data analysis approach. Section 2.8.3 introduces models with nonparametric functions m_l.

2.8.1 Dynamic Parametric Factor Models

The simplest approach to adopt (2.40) is to assume some parametric structure for functions m_l. While the functions are given the estimation of Z_{tl} simplifies to the parametric regression problem from Section 2.3 applied for each t separately. Note that the vector $(m_0(X_{tj}), \ldots, m_L(X_t j))^\top$ can be considered as a new external exploratory variable and whole analysis from Section 2.6 may be applied.

For the illustration purpose we present an example from Nelson and Siegel (1987) on dynamics of the term structure. The model has a form

$$E(Y_t|X_t) = Z_{t1} + Z_{t2}\frac{1-e^{-\lambda X_t}}{\lambda X_t} - Z_{t3}\left(\frac{1-e^{-\lambda X_t}}{\lambda X_t} - e^{-\lambda X_t}\right), \qquad (2.41)$$

where Y_t describes yields of zero coupon bonds at time t and X_t the corresponding maturities. The functions m_l are easy to read from (2.41). Note that they contain also one embedded parameter. Generally speaking, it can be obtained by certain modelling experience or interpretation arguments e.g. one imposes an orthogonality constrain or optimize for the best fit of the data. However, until the number of parameters are fixed and not free to increase, the modelling framework remains in the parametric world. Allowing the number of parameter to increase leads to nonparametric and semiparametric models. These cases are discussed later on in Section 2.8.2 and Section 2.8.3.

The model (2.41) simplifies the dynamics of complex functional object to three factors. The inference on the whole curve reduces to the inference on the time varying factors. Diebold and Li (2006) interpret them as level, slope and curvature, and they propose the autoregressive process for studying the forecasting power of the model. Diebold et al. (2006) include the macroeconomic variables and find an evidence of interrelations between macro variables and future movements in the yield curve.

As a second example we consider the model of a form

$$Y_{tj} = Z_{t0} + \sum_{i=1}^{d} X_{tj}^{(i)} Z^{(i)} + \varepsilon_{tj}. \qquad (2.42)$$

This model is a special case of (2.40) with two parametric functions, where $Z = (Z^{(1)}, \ldots, Z^{(d)})^\top$ is d-dimensional parameter. The whole dynamics is then described by one dimensional time series Z_{t0}.

Note that (2.42) is a linear regression with common slope and time varying intercept. It has a similar structure to (2.34) with $V_j = 0$. In this case time

effects in panel data may be interchanged with individual effects and the model simplifies to (2.22) with swapped indexes j and t. In order to estimate Z_{t0} and Z one has to proceed like in Section 2.6.1. Note that appropriate adjustment for unbalanced panel is required but the estimators $\widehat{Z}_{t0}$ and $\widehat{Z}$ remain mainly in the form of (2.26) and (2.25).

The model is applied by Hansen et al. (2004) to risk theory for prediction of growth of claim sizes. The paper presents some theoretical results on statistical analysis of the time series. Since Z_{t0} is not observed one has only $\widehat{Z}_{t0}$ at hand, which can be estimated accurately and the estimated time series can be analyzed as the time series itself. Under some regularity conditions, for details see Hansen et al. (2004),

$$\sum_{t=1}^{T}(\widehat{Z}_{t0} - Z_{t0})^2 = o_P(1). \tag{2.43}$$

Assuming now AR(1) process for Z_{t0} and $\widehat{Z}_{t0}$ the statistical properties of the parameter's estimator are similar to what one would obtain were Z_{t0} given.

2.8.2 Functional Data Analysis

In Section 2.7 we have discussed functional data analysis. In particular the fPCA, the Karhunen-Loève expansion and estimation methods are presented. Here, we recall this methodology and based on the best empirical basis property show how the functional data analysis can serve as a factorial method in the form of (2.40).

The first step is to obtain the estimates $\widehat{Y}_t(x)$ on a grid or by the series estimator. Some methods from Section 2.3 can be applied. Note that in this pre-estimation phase T separate regressions have to be run. The main aim is to regularize the data in order to proceed later on in more convenient multivariate context.

The implicit assumption of this procedure states that initial approximation at time t is accurate. Therefore the error of shifting the analysis from the observed data to their regular equivalent should be negligible. Note that this initial estimation step keeps this method in nonparametric context since up to the certain level the grid or series expansion can contain any number of points or basis functions respectively.

In the second step one proceeds like in Section 2.7.3. Then the estimates of $\widehat{PC}_t$ and $\widehat{\gamma}$ are obtained and assigned to $\widehat{Z}_t$ and $(\widehat{m}_1, \ldots, \widehat{m}_L)^\top$ respectively. The functions $\widehat{m}_l(x)$ have also representation on a grid or by a series expansion. Contrary to the parametric approach they are not given in advance and their form depends on the data.

Since the method performs basically in multivariate context only N (number of grid points) or K (number of series functions) components are localized. However, it is still a significant number and to obtain a low dimensional representation the full model has to be cut down to L first functions. For this purpose some ideas presented in Section 2.4.2 may be employed. Note that the model is nested so increasing L by one has no influence on the previous components.

The PCA approach gained great popularity in financial modelling of dynamic phenomena. Rebonato (1998), Bliss (1997), Molgedey and Galic (2001) among others applied the PCA to study the yield curves. Cont and Fonseca (2002) proposed it for modelling of dynamics of implied volatility surfaces. Koekebakker and Ollmar (2005) studied by PCA the term structure of electricity futures prices.

2.8.3 Dynamic Semiparametric Factor Models

While the models in Section 2.8.1 assume parametric form of the loading functions m_l, Section 2.8.2 presents approach with functions estimated from the data itself. In this part we keep nonparametric m_l but simultaneously impose some structure on the model, namely we fix the number of factors L in advance. Therefore, we refer to this modelling strategy as *semiparametric* since it joins the nonparametric flexibility with enforced structure.

The main difference to approach from Section 2.8.1 is the lack of parametric form of the functions. In contrast to functional data approach no initial estimation of Y_t is necessary and the estimators of factors comes directly from the data. However, the models from this section are not nested since the number of factors have to be specified.

An instructive example is a famous model from Lee and Carter (1992). It serves as a tool for the mortality forecast while it models the age specific death rates as a sum of time invariant age specific component, and a product of another age specific component and time varying factor. More precisely, it can be written

$$Y_{tj} = m_0(X_{tj}) + Z_{t1}m_1(X_{tj}) + \varepsilon_{tj},$$

where Y_{tj} is logarithm of the age specific death rates observed at time t for the group j and X_{tj} the corresponding age. Y_t and X_t can be regarded as a $(J \times 1)$ vectors and after estimating m_0 as a sample mean the methodology from Section 2.4.2 or Section 2.8.2 can be employed to obtain $\widehat{Z}_{t1}$ and $\widehat{m}_1$. Due to some identification issues, discussed later on, $\widehat{Z}_{t1}$ can be normalized to have zero mean. The model is relatively simple and the forecast can be

obtained by statistical analysis of the one dimensional time series. In Lee and Carter (1992)the model is fitted to the U.S. mortality data 1900-1989 and mortality forecast is obtained through random walk with drift.

Another example comes from Fama and French (1992), who proposed a factor model for stock returns. Here we discuss the semiparametric extension studied by Connor and Linton (2007) or Connor et al. (2007). The stock returns Y_{tj} are regressed on the asset characteristics X_{tj} in the following way

$$Y_{tj} = Z_{t0} + \sum_{l=1}^{L} Z_{tl} m_l(X_{tj}^{(l)}) + \varepsilon_{tj}, \tag{2.44}$$

where m_l are univariate functions of the lth characteristic's coordinate. Z_{t0} is the common trend and Z_{tl} for $l = 1, \dots L$ are common factors loaded with functions m_l. Similarly to factors models from Section 2.4 ε_{tj} denotes the asset specific returns.

The model resembles the additive models from Section 2.3.3 with time varying weight. For identifying the model one may assume $E\{m_l(X_{tj}^{(l)})\} = 0$ and $Var\{m_l(X_{tj}^{(l)})\} = 1$.

The estimation algorithm, described in details in Connor et al. (2007), utilizes kernel based approach and searches for $(\widehat{Z}_t, \widehat{m})$ minimizing

$$\frac{1}{T}\sum_{t=1}^{T} E\left[\left\{Y_{tj} - Z_{t0} - \sum_{l=1}^{L} Z_{tl} m_l(X_{tj}^{(l)})\right\}^2\right],$$

where the expectation is taken over returns and characteristics. It reflects some similarities to the backfitting algorithm, see Section 2.3.3, and share the same modelling perspective to the method discussed later in Section 2.9.1.

The procedure works iteratively by constant adjusting between $\widehat{Z}_{tl}$ and $\widehat{m}$. Note that given rth step estimator updating $\widehat{Z}_t$ reduces to the least square problem solved separately for each t. The adjustment for $\widehat{m}$ in point x is obtained from

$$m_{l'}(x) = \frac{\sum_{t,j} \frac{Z_{tl'}}{J_t}\left\{E(Y_{tj}|X_{tj}^{(l')} = x) - Z_{t0} - \sum_{l \neq l'} Z_{tl} E(m_l(X_{tj}^{(l)})|X_{tj}^{(l')} = x)\right\}}{\sum_{t=1}^{T} Z_{tl'}^2}$$

by plugging in the right hand side appropriate estimators. For estimating conditional expected value the Nadaraya-Watson estimator could be used. Hence

$$\widehat{E}_t\left\{Y_{tj}|X_{tj}^{(l)} = x\right\} = \frac{\sum_{j=1}^{J_t} \mathcal{K}_h(X_{tj}^{(l)} - x) Y_{tj}}{\sum_{j=1}^{J_t} \mathcal{K}_h(X_{tj}^{(l)} - x)}.$$

Similarly for the second expectation

$$\widehat{E}_t(m_l(X_{tj}^{(l)})|X_{tj}^{(l)} = x) = \frac{\sum_{j=1}^{J_t} \mathcal{K}_h(X_{tj}^{(l)} - x)\widehat{m}_l(X_{tj}^{(l)})}{\sum_{j=1}^{J_t} \mathcal{K}_h(X_{tj}^{(l)} - x)}. \tag{2.45}$$

Note that expectation estimators are time dependent, which is indicated by the subindex t. Additionally in (2.45) one has to plug some estimator $\widehat{m}_l$. In the iterative procedure it is usually last given estimator. As a consequence the $(r+1)$ update of the $\widehat{m}_{l'}$ has a form

$$\begin{aligned}\widehat{m}_{l'}^{(r+1)}(x) = &\frac{\sum_{t=1}^{T} \widehat{Z}_{tl'}^{(r)}(\widehat{E}_t(Y_{tj}|X_{tj}^{(l')} = x) - \widehat{Z}_{t0}^{(r)})}{\sum_{t=1}^{T} \widehat{Z}_{tl'}^{(r)2}} \\ &- \frac{\sum_{t=1}^{T} \sum_{l>l'} \widehat{Z}_{tl'}^{(r)} \widehat{Z}_{tl}^{(r)} \widehat{E}_t^{(r)}}{\sum_{t=1}^{T} \widehat{Z}_{tl'}^{(r)2}} \\ &- \frac{\sum_{t=1}^{T} \sum_{l<l'} \widehat{Z}_{tl'}^{(r)} \widehat{Z}_{tl}^{(r)} \widehat{E}_t^{(r+1)}}{\sum_{t=1}^{T} \widehat{Z}_{tl'}^{(r)2}},\end{aligned}$$

where $\widehat{E}_t^{(r)}$ denotes the abbreviated term of (2.45) with r-th update plugged in for the $\widehat{m}_l$ estimator. Note that similarly to the backfitting algorithm the last obtained estimator is used for next iteration.

The iteration runs until convergence criteria are met. The identification restrictions can be obtained by replacing the final estimators of the $\widehat{m}_l$ with their standardized equivalents. In order to obtain starting values $\widehat{m}_l^{(0)}$ of the algorithm one may run additive model on the pooled data or assume some parametric form e.g. identity. The second approach is equivalent to use some methods from panel data analysis.

The two presented examples represent semiparametric approach for modelling dynamic regression problems. They impose, however, some additional limitations. Lee-Carter model applied to relatively regular data design and in fact can be seen as some special case of the classical multivariate statistical analysis. The semiaparmetric model for stock returns extends the panel modelling. Since no asset's specifics except X_{tj} are included in the model, we find it justified to present it rather in this place. The functions m_l, however, are univariate, which gives the closer connections to additive models than the factor analysis from Section 2.4.

In this work we focus on general semiparametric representation of (2.40). To be more specific when we refer to the dynamic semiparametric factor model (DSFM) we mean the model of the form

$$E(Y_t|X_t) = m_0(X_t) + \sum_{l=1}^{L} Z_{tl} m_l(X_t). \tag{2.46}$$

Here we impose that $Z_{t0} = 1$ and $m_l : \mathbb{R}^d \mapsto \mathbb{R}$. Even if X_t is one dimensional one can enforce $L > 1$ which is inline with the factor models from Section 2.4. In (2.44) there exist some common dynamic trend Z_{t0}. The factors and loadings (Z_t, m_l) for $l = 1, \ldots, L$ describe the discrepancy between the trend and the particular object. In (2.46) by setting $Z_{t0} = 1$ the time varying factors reflect the distortion from the common level given by m_0.

The model can be applied in the analysis of the term structure dynamics. Apart from aforementioned yield curves one may be interested in the term structure of variance swaps, see Detlefsen and Härdle (2006), or futures prices of CO_2 emission allowance, see Trück et al. (2006). However the most prominent example is the dynamics of implied volatility surface, studied with the DSFM by Fengler et al. (2007), Borak et al. (2005), Borak et al. (2007). The detailed discussion on the applications of the DSFM is given in Chapter 4.

2.9 Dynamic Semiparametric Factor Models

In Section 2.8.3 we only shortly present the DSFM. More detailed study on the model's properties is left to this section. In Section 2.9.1 we present the estimation algorithms, the asymptotic behavior is given in Section 2.9.2 and the model selection issues are considered in Section 2.9.3.

In previous sections we present some modelling strategies. The presentation is focussed on the relation to the DSFM. Before we discuss the properties of the model we briefly summarize the statistical modelling techniques by pointing out their connections to the DSFM.

In our short excursion on the modelling techniques we use Figure 2.2, which provide an intuitive illustration. The upper left corner shows the classical time series case. One observes the evolution of several objects through time without any relationship between them. As an example consider the measurements of macroeconomic variables like unemployment, inflation rate, gross domestic product or collection of the stock prices. When the dimensionality of the time series is high some factor analysis can be employed in order to make the modelling feasible. This case is illustrated in upper right panel, which depicts some relation between true series and factors. The dynamics of the observed time series is now approximated by the dynamics of the factors.

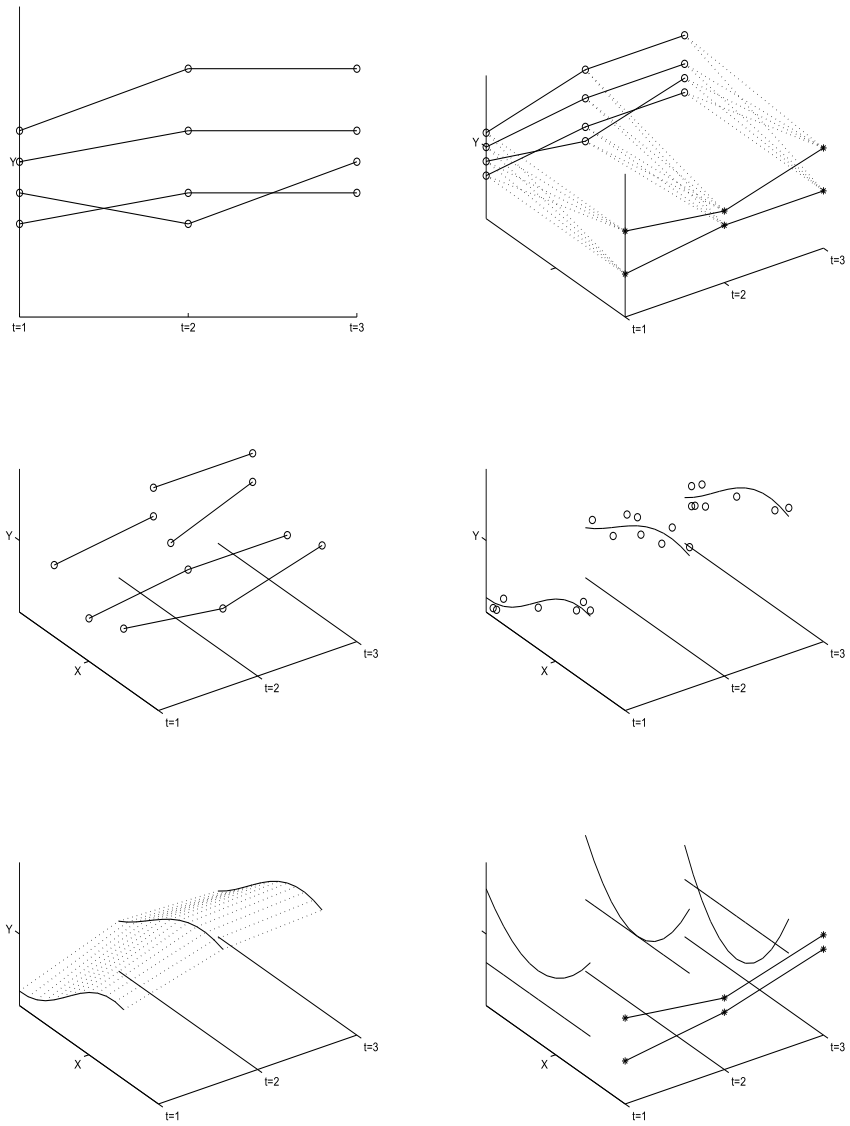

Figure 2.2: *An illustrative example for some modelling techniques. Upper panels present time series models and application of factor techniques to the time series modelling. Middle left panel represents unbalanced panel data and the middle right one regression problems evolving in time. Lower left panel shows discrete representation of the regression curves. Lower right panel illustrates the factor models idea for the dynamic regression problems.*

The middle panel present situations where one has exploratory variables at hand. The left one corresponds to the panel data, in particular unbalanced panel. One observes the evolution of some individuals through time, possibly not each object on the whole time range, together with their characteristics, which serve as some kind of linkage, e.g. record of a stock. This connection in the exploratory space is implicit in contrast to the direct connection in time space since one simply observes the evolution of the same entity. This feature is reflected in the modelling by inclusion of the object specifics in form of fixed or random effects. However, when there is no link in time like in implied volatility intraday data, see Section 4.1.1, one has a regression problem, developing through time. The middle right panel illustrate this situation. In each time point there could be a different number of observations with completely different design.

The common modelling technique for this phenomena utilize the factor approach, which is visualized in the right bottom panel. The target regression curve is approximated through a weighted sum of some specific functions and its dynamics obtained by allowing the weights to vary in time. The time invariant functions can be estimated with parametric or nonparametric methods. In the parametric approach one has to usually solve sequence of the least square problems. In the nonparametric framework on possible way to proceed is to find finite approximation of the regression curve on each time period separately, see left bottom panel, and apply some standard techniques to the approximation points, treating them as if they were observed. The methods, referred here as semiparametric, avoid this step and obtain the functions and factors directly from the data.

Note that the methods for analyzing dynamic regression, like the DSFM, can be also applied on the panel data. In this case, however, one omits the information given in the time evolution of the individuals. It could be exemplified by the method of Connor and Linton (2007). For the time series defining the regressors as a lagged observation of the component series yield an alternative dimension reduction technique. In particular the DSFM can be regarded as nonparametric time series method.

2.9.1 Estimation

This part is devoted to the estimation algorithms of the DSFM. We present mainly two approaches which have been already mentioned namely: representation on the discrete grid and series estimation.

Before we discuss in details the estimating procedures we recall once more the structure of the model. Let (X_{tj}, Y_{tj}) be observed for $j = 1, \ldots, J_t$ and $t = 1, \ldots, T$. Then

$$Y_{tj} = m_0(X_{tj}) + \sum_{l=1}^{L} Z_{tl} m_l(X_{tj}) + \varepsilon_{t,j} = Z_t^\top m(X_{tj}) + \varepsilon_{tj}, \tag{2.47}$$

where $Z_t = (Z_{t0}, \ldots, Z_{tL})^\top$ is an unobservable $(L+1)$-dimensional process with $Z_{t0} \equiv 1$ and the function m is a tuple $(m_0, \ldots, m_L)$ of unknown real-valued functions m_j defined on a subset of $\mathbb{R}^d$. The variables $X_{11}, \ldots, X_{TJ_T}$, $\varepsilon_{11}, \ldots, \varepsilon_{TJ_T}$ are independent. The errors ε_{tj} have zero means and finite second moments. We assume that the covariates X_{tj} have support $\Xi = [0,1]^d$ and for simplicity of notation in the series estimators part we set that $J_t \equiv J$ do not depend on t.

Kernel Approach

First we present the kernel type estimator proposed by Fengler et al. (2007) and studied also by Borak et al. (2005). The estimates $\widehat{Z}_{tl}$ and $\widehat{m}_l$ are obtained by minimizing the following least squares criterion

$$\sum_{t=1}^{T} \sum_{j=1}^{J_t} \int \left\{ Y_{tj} - \sum_{l=0}^{L} \widehat{Z}_{tl} \widehat{m}_l(u) \right\}^2 \mathcal{K}_h(u - X_{tj})\, du, \tag{2.48}$$

where $\mathcal{K}_h$ denotes the kernel function. The minimization procedure search through all functions $\widehat{m}_l : \Xi \longrightarrow \mathbb{R}$ $(l = 0, ..., L)$ and time series $\widehat{Z}_{tl} \in \mathbb{R}$. Note that for $L = 0$ the procedure reduces to Nadaraya-Watson estimator based on pooled sample of all time periods, see Section 2.3.2.

To calculate the estimates an iterative procedure is applied. We introduce the following notation for $1 \le t \le T$

$$\widehat{p}_t(u) = \frac{1}{J_t} \sum_{j=1}^{J_t} \mathcal{K}_h(u - X_{tj}), \tag{2.49}$$

$$\widehat{q}_t(u) = \frac{1}{J_t} \sum_{j=1}^{J_t} \mathcal{K}_h(u - X_{tj}) Y_{tj}. \tag{2.50}$$

We denote by $\widehat{m}^{(r)} = (\widehat{m}_0^{(r)}, ..., \widehat{m}_L^{(r)})^\top$ and $\widehat{Z}_t^{(r)} = (\widehat{Z}_{tl}^{(r)}, ..., \widehat{Z}_{tL}^{(r)})^\top$ the estimates after r iterations. By replacing each function $\widehat{m}_l$ in (2.48) by $\widehat{m}_l + \delta g$ with arbitrary function g and taking derivatives with respect to δ one obtains

$$\frac{d}{d\delta} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \int \left\{ Y_{tj} - \sum_{l=0}^{L} \widehat{Z}_{tl} \widehat{m}_l(u) - \widehat{Z}_{tl'} \delta g(u) \right\}^2 \mathcal{K}_h(u - X_{tj})\, du =$$

$$2\int \sum_{t=1}^{T}\sum_{j=1}^{J_t}\left\{Y_{tj}-\sum_{l=0}^{L}\widehat{Z}_{tl}\widehat{m}_l(u)-\widehat{Z}_{tl'}\delta g(u)\right\}\widehat{Z}_{tl'}g(u)\mathcal{K}_h(u-X_{tj})\,du=0. \tag{2.51}$$

Since the minimum is obtained for $\delta = 0$ and for any function g the integral in (2.51) is 0 if

$$\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left\{Y_{tj}-\sum_{l=0}^{L}\widehat{Z}_{tl}\widehat{m}_l(u)\right\}\widehat{Z}_{tl'}\mathcal{K}_h(u-X_{tj})=0. \tag{2.52}$$

Rearranging terms in (2.52) and plugging in (2.49)-(2.50) yields

$$\sum_{t=1}^{T}J_t\widehat{Z}_{tl'}\widehat{q}_t(u)=\sum_{t=1}^{T}J_t\sum_{l=0}^{L}\widehat{Z}_{tl'}\widehat{Z}_{tl}\widehat{p}_t(u)\widehat{m}_l(u), \tag{2.53}$$

for $0 \le l' \le L$. In fact (2.53) is a set of $L+1$ equations. Define the matrix $B^{(r)}(u)$ and vector $Q^{(r)}(u)$ by their elements:

$$\left(B^{(r)}(u)\right)_{l,l'}=\sum_{t=1}^{T}J_t\widehat{Z}_{tl'}^{(r-1)}\widehat{Z}_{tl}^{(r-1)}\widehat{p}_t(u), \tag{2.54}$$

$$\left(Q^{(r)}(u)\right)_{l}=\sum_{t=1}^{T}J_t\widehat{Z}_{tl}^{(r-1)}\widehat{q}_t(u). \tag{2.55}$$

Thus (2.53) is equivalent to

$$B^{(r)}(u)\widehat{m}^{(r)}(u)=Q^{(r)}(u), \tag{2.56}$$

which yields the estimate of $\widehat{m}^{(r)}(u)$ in the rth iteration. The estimate is given on a particular point u and for practical usage the functions are represented on the grid.

A similar idea has to be applied to update $\widehat{Z}_t^{(r)}$. Replacing $\widehat{Z}_{tl}$ by $\widehat{Z}_{tl}+\delta$ in (2.48) and taking once more the derivative with respect to δ yield

$$\sum_{j=1}^{J_t}\int\left\{Y_{tj}-\sum_{l=0}^{L}\widehat{Z}_{tl}\widehat{m}_l(u)\right\}\widehat{m}_{l'}(u)\mathcal{K}_h(u-X_{tj})du=0,$$

which leads to

$$\int\widehat{q}_t(u)\widehat{m}_{l'}(u)\,du=\sum_{l=0}^{L}\widehat{Z}_{tl}\int\widehat{p}_t(u)\widehat{m}_{l'}(u)\widehat{m}_l(u)\,du, \tag{2.57}$$

for $1 \leq l' \leq L$. The formula (2.57) is now a system of L equations. Define the matrix $M^{(r)}(i)$ and the vector $S^{(r)}(i)$ by their elements

$$\left(M^{(r)}(t)\right)_{l,l'} = \int \widehat{p}_t(u)\widehat{m}_{l'}(u)\widehat{m}_l(u)\ du,$$

$$\left(S^{(r)}(t)\right)_l = \int \widehat{q}_t(u)\widehat{m}_l(u)\ du - \int \widehat{p}_t(u)\widehat{m}_0(u)\widehat{m}_l(u)\ du.$$

An estimate of $\widehat{Z}_t^{(r)}$ is thus given by solving

$$M^{(r)}(t)\widehat{Z}_t^{(r)} = S^{(r)}(t). \tag{2.58}$$

The algorithm stops when only minor changes occur

$$\sum_{i=t}^{T} \int \left(\sum_{l=0}^{L} \widehat{Z}_{tl}^{(r)} \widehat{m}_l^{(r)}(u) - \widehat{Z}_{tl}^{(r-1)} \widehat{m}_l^{(r-1)}(u) \right)^2 du \leq \epsilon$$

for some small ϵ.

The successful estimation may suffer from the unequal distribution of the data points. Gaps in some particular design range may obstruct the estimation procedure. If in any point u' the function $\widehat{p}(u') = 0$ then obviously matrix $B^{(r)}(u')$ in (2.54) contains only 0 and is singular. This means that one may not estimate successfully any value in the local neighborhood of u'. This problem may be solved by increasing bandwidths, considering k-nearest neighbor estimator or local bandwidths. The problem of the gap in the data can also be partially handled by the choice of initial estimates $\widehat{Z}_t^{(0)}$. In Fengler et al. (2007) a piecewise constant initial time series are proposed. The subintervals $\mathcal{T}_1, ..., \mathcal{T}_L$ are pairwise disjoint subsets of $\{1, ..., T\}$ and $\bigcup_{l=1}^{L} \mathcal{T}_l$ is a strict subset of $\{1, ..., T\}$. The initial estimates are now defined by $\widehat{Z}_{tl}^{(0)} = 1$ if $t \in \mathcal{T}_l$ and $\widehat{Z}_{tl}^{(0)} = 0$ if $i \notin \mathcal{T}_l$. To complete the setting $\widehat{Z}_{t0}^{(0)} = 1$ for each t. However, this kind of setting requires even more data to obtain the final estimates. For each subset $\mathcal{T}_l$ there need to exist at least one day $t \in \mathcal{T}_l$ such that $\widehat{p}_t(u') \neq 0$, otherwise the row of zeros in (2.54) appears. The smaller is the length of $\mathcal{T}_l$ intervals the bigger bandwidths need to be taken. This deficiency can be removed by taking a random initial time series. Then $\widehat{p}_t(u)$ needs to be non-zero for one t in $\{1, ..., T\}$ and it is no longer necessary that $\widehat{p}_t(u)$ is non-zero for one t in each $\mathcal{T}_l$. The empirical illustration of this issue is given in Borak et al. (2005).

Series Estimators

Here for the estimation of m, we use a series estimator. For an integer $K \geq 1$, we choose functions $\psi_1, \ldots, \psi_K : [0,1]^d \to \mathbb{R}$ which are normed so

that $\int_{[0,1]^d} \psi_k^2(x)\,dx = 1$. For example, one may take $\{\psi_k : 1 \le k \le K\}$ to be a B-spline basis, see Section 2.1. Then, a tuple of functions $m = (m_0, \ldots, m_L)^\top$ may be approximated by $A\psi$, where $A = (\alpha_{l,k})$ is a $(L+1)\times K$ matrix and $\psi = (\psi_1, \ldots, \psi_K)^\top$. We define the least squares estimators $\widehat{Z}_t = (\widehat{Z}_{t,0}, \ldots, \widehat{Z}_{t,L})^\top$ and $\widehat{A} = (\widehat{\alpha}_{l,k})$

$$S(A, Z) \equiv \sum_{t=1}^{T} \sum_{j=1}^{J} \left\{ Y_{tj} - Z_t^\top A \psi(X_{tj}) \right\}^2 = \min_{A,Z}! \tag{2.59}$$

where $Z = \left(Z_1^\top, \ldots, Z_T^\top\right)^\top$. The minimization runs over all values of $\widehat{Z}_t$ with

$$Z_{t0} = 1. \tag{2.60}$$

With $\widehat{A}$ at hand, we estimate m by $\widehat{m} = \widehat{A}\psi$.

We note that, given Z or A, the function S in (2.59) is quadratic with respect to the other variables, and thus has an explicit unique minimizer. However, minimization of L with respect to A and Z simultaneously is a fourth-order problem. The solution is neither unique nor explicit. It is unique only up to the values of $\widehat{Z}_1^\top \widehat{A}, \ldots, \widehat{Z}_T^\top \widehat{A}$. We will come back to this identifiability issue later in this section.

Before we present the estimation algorithm we first introduce some additional matrix notation. Let $\alpha \equiv \alpha(A)$ denote the stack form of $A = (\alpha_{l,k})$, i.e.,

$$\alpha = (\alpha_{0,1}, \ldots, \alpha_{L,1}, \alpha_{0,2}, \ldots, \alpha_{L,2}, \ldots, \alpha_{0,K}, \ldots, \alpha_{L,K})^\top.$$

Let $\Psi_t = \left[\psi(X_{t1}), \ldots, \psi(X_{tJ})\right]$ be a $(K \times J)$ matrix, $Y_t = (Y_{11}, \ldots, Y_{1J})^\top$ and $Y = (Y_1, \ldots, Y_T)^\top$ a $(JT \times 1)$ vector.

To find a solution $(\widehat{A}, \widehat{Z})$ of the minimization problem (2.59), one might adopt the following iterative algorithm:
(i) Given an initial choice $Z^{(0)}$, minimize $S(A, Z^{(0)})$ with respect to A, which is an ordinary least squares problem and thus has an explicit unique solution. Call it $A^{(1)}$ and its corresponding stack form $\alpha^{(1)}$. Define $\Psi_Z = \Big(\Psi_1 \otimes Z_1, \ldots, \Psi_T \otimes Z_T\Big)^\top$ then

$$\alpha^{(1)} = \left(\Psi_Z^\top \Psi_Z\right)^{-1} \Psi_Z^\top Y. \tag{2.61}$$

(ii) Minimize $S(A^{(1)}, Z)$ with respect to Z. This leads also to ordinary least squares problems. Note that given A the update $Z_t^{(1)}$ can be obtained separately for each t, i.e.

$$Z_{t,2:L}^{(1)} = \left(A_{2:L} \Psi_t \Psi_t^\top A_{2:L}^\top\right)^{-1} A \Psi_t (Y_t - \Psi_t^\top A_{1:1}^\top), \tag{2.62}$$

where $Z_{t,2:L}$ and $A_{2:L}$ denote the corresponding exclusion of the first row, and $A_{1:1}$ is the first row of matrix A. This additional complication comes from the constraint (2.60).
(iii) Iterate (i) and (ii) until convergence.

This approach is very much similar to the method presented in the kernel estimation part. However, the procedure is not guaranteed to converge to a solution of the original problem.

In addition to the presented procedure we consider as well a Newton-Raphson algorithm. In a slight abuse of notation we write $S(\alpha, Z)$ for $S(A, Z)$. Define

$$\begin{aligned} F_{10}(\alpha, Z) &= \frac{\partial}{\partial \alpha} S(\alpha, Z), \quad F_{01}(\alpha, Z) = \frac{\partial}{\partial Z} S(\alpha, Z), \\ F_{20}(\alpha, Z) &= \frac{\partial^2}{\partial \alpha^2} S(\alpha, Z), \quad F_{02}(\alpha, Z) = \frac{\partial^2}{\partial Z^2} S(\alpha, Z), \\ F_{11}(\alpha, Z) &= \frac{\partial^2}{\partial \alpha \partial Z} S(\alpha, Z). \end{aligned}$$

Then, it can be shown that

$$\begin{aligned} F_{10}(\alpha, Z) &= 2 \sum_{t=1}^{T} \left[(\Psi_t \Psi_t^\top) \otimes (Z_t Z_t^\top) \right] \alpha - 2 \sum_{t=1}^{T} (\Psi_t Y_t) \otimes Z_t, \\ F_{20}(\alpha, Z) &= 2 \sum_{t=1}^{T} \left[(\Psi_t \Psi_t^\top) \otimes (Z_t Z_t^\top) \right], \\ F_{01}(\alpha, Z) &= 2 \begin{pmatrix} A\Psi_1 \Psi_1^\top A^\top Z_1 - A\Psi_1 Y_1 \\ A\Psi_2 \Psi_2^\top A^\top Z_2 - A\Psi_2 Y_2 \\ \vdots \\ A\Psi_T \Psi_T^\top A^\top Z_T - A\Psi_T Y_T \end{pmatrix}, \\ F_{02}(\alpha, Z) &= 2 \begin{pmatrix} A\Psi_1 \Psi_1^\top A^\top & 0 & \cdots & 0 \\ 0 & A\Psi_2 \Psi_2^\top A^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & A\Psi_T \Psi_T^\top A^\top \end{pmatrix}. \end{aligned}$$

By some algebraic manipulations it can be shown that

$$\left[(\Psi_t \Psi_t^\top) \otimes (Z_t Z_t^\top) \right] \alpha = (\Psi_t \Psi_t^\top A^\top Z_t) \otimes Z_t.$$

Thus, we get

$$F_{11}(\alpha, Z) = 2 \left(F_{11,1}(\alpha, Z),\ F_{11,2}(\alpha, Z),\ \ldots,\ F_{11,T}(\alpha, Z) \right),$$

where

$$F_{11,t}(\alpha, Z) = (\Psi_t \Psi_t^\top A^\top) \otimes Z_t + (\Psi_t \Psi_t^\top A^\top Z_t) \otimes I_{L+1} - (\Psi_t Y_t) \otimes I_{L+1}$$

and I_q denotes the identity matrix of dimension q. Let

$$F(\alpha, Z) = \begin{pmatrix} F_{10}(\alpha, Z) \\ F_{01}(\alpha, Z) \end{pmatrix}, \quad F'(\alpha, Z) = \begin{pmatrix} F_{20}(\alpha, Z) & F_{11}(\alpha, Z) \\ F_{11}(\alpha, Z)^\top & F_{02}(\alpha, Z) \end{pmatrix}.$$

We need to solve the equation $F(\alpha, Z) = 0$ simultaneously for α and Z. Given (α^{OLD}, Z^{OLD}), the Newton-Raphson algorithm gives the updating equation for (α^{NEW}, Z^{NEW})

$$\begin{pmatrix} \alpha^{NEW} \\ Z^{NEW} \end{pmatrix} = \begin{pmatrix} \alpha^{OLD} \\ Z^{OLD} \end{pmatrix} - F'(\alpha^{OLD}, Z^{OLD})^{-1} F(\alpha^{OLD}, Z^{OLD}). \quad (2.63)$$

The algorithm (2.63) is shown to converge to a solution of (2.59) at a geometric rate under some week conditions on the initial choice $(\alpha^{(0)}, Z^{(0)})$, as is demonstrated by Theorem 2.9.1 below. We collect the conditions for the theorem.

(C1) $F'(\alpha^{(0)}, Z^{(0)})$ is invertible.

(C2) There exists a solution $(\widehat{\alpha}, \widehat{Z})$ of the equation $F(\alpha, Z) = 0$ such that $\sum_{t=1} \widehat{Z}_t \widehat{Z}_t^\top$ and $\sum_{t=1} \widehat{Z}_t Z_t^{(0)\top}$ are invertible, and $\widehat{\alpha}_l = (\widehat{\alpha}_{l1}, \ldots, \widehat{\alpha}_{lK})^\top$ for $l = 0, \ldots, L$ are linearly independent, i.e., the matrix $\widehat{A}$ that corresponds to $\widehat{\alpha}$ has full rank.

Let $\alpha^{(k)}$ and $Z^{(k)}$ denote the kth updated vectors in the iteration with the algorithm (2.63). Also, we write $A^{(k)}$ for the matrix that corresponds to $\alpha^{(k)}$.

Theorem 2.9.1 *Suppose that the initial choice $(\alpha^{(0)}, Z^{(0)})$ satisfies (C1) and (C2). Then, for any $0 < \gamma < 1$ there exists $r > 0$ and $C > 0$ such that, if $\sum_{t=1}^T \| \widetilde{Z}_t^{(0)\top} A^{(0)} - \widehat{Z}_t^\top \hat{A} \|^2 \leq r$, then*

$$\sum_{t=1}^T \| Z_t^{(k)\top} A^{(k)} - \widehat{Z}_t^\top \hat{A} \|^2 \leq C 2^{-(k-1)} \gamma^{2^k - 1}.$$

The minimization problem (2.59) has no unique solution. If $(\widehat{Z}_t, \widehat{A})$ or $(\widehat{Z}_t, \widehat{m} = \widehat{A}\psi)$ is a minimizer, then also $(\widetilde{B}^\top \widehat{Z}_t, \widetilde{B}^{-1} \widehat{m})$ is a minimizer. Here $\widetilde{B}$ is an arbitrary matrix of the form

$$\widetilde{B} = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} \quad (2.64)$$

for an invertible matrix B. The special structure of $\widetilde{B}$ assures that the first component of $\widetilde{B}^\top \widehat{Z}_t$ equals 1.

Below we present the proposition that show that the algorithm (2.63) is independent on the identification issues. Let (α^{OLD}, Z^{OLD}) be the 'old' value to be updated by (2.63) and (α^{NEW}, Z^{NEW}) be the updated value as defined there. For an $(L+1) \times (L+1)$ nonsingular square matrix $\widetilde{B}$, define

$$\widetilde{A}^{OLD} = \widetilde{B}^{-1} A^{OLD}, \quad \widetilde{Z}_t^{OLD} = \widetilde{B}^\top Z_t^{OLD}.$$

The stack form of $\widetilde{A}^{OLD}$ can be written as

$$\widetilde{a}^{OLD} = (I_K \otimes \widetilde{B}^{-1}) \alpha^{OLD}. \tag{2.65}$$

Also, we have

$$\widetilde{Z}^{OLD} = (\widetilde{Z}_1^{OLD\top}, \ldots, \widetilde{Z}_T^{OLD\top})^\top = (I_T \otimes \widetilde{B}^\top) Z^{OLD}. \tag{2.66}$$

The proposition demonstrates that updating $\widetilde{a}^{OLD}$ and $\widetilde{Z}^{OLD}$ directly by the formula (2.63) is equivalent to first updating α^{OLD} and Z^{OLD} by (2.63) and then transforming the updated vectors α^{NEW} and Z^{NEW} according to (2.65) and (2.66), respectively.

Proposition 2.9.2 *Let $\widetilde{a}^{NEW}$ and $\widetilde{Z}^{NEW}$ be obtained from $\widetilde{a}^{OLD}$ and $\widetilde{Z}^{OLD}$ by the updating equation (2.63). Then $\widetilde{a}^{NEW} = (I_K \otimes \widetilde{B}^{-1}) \alpha^{NEW}$ and $\widetilde{Z}^{NEW} = (I_T \otimes \widetilde{B}^\top) Z^{NEW}$.*

The proof of this Proposition is given in Borak et al. (2007).

Numerical Analysis

In this part we briefly discuss the numerical issues concerning the presented algorithms for the DSFM estimation. As it has been already indicated the methods contain the series of least square problems. For the simplicity of our analysis we assume that standard matrix inversion with dimension n is $\mathcal{O}(n^3)$, although methods with lower computational complexity have been already proposed, see Coppersmith and Winograd (1990).

We start our discussion with the kernel method. The functions $\widehat{m}_l$ are given on a grid. Let this grid contain N points. First (2.49) and (2.50) have to be obtained which is in general $\mathcal{O}(TJN)$. Note that this evaluation of the kernel functions appears only in the first stage of the algorithm and does not need to be recalculated in the iteration stage. In the each circle one has to solve (2.56) for each grid poind. This leads to $\mathcal{O}(N(L+1)^3)$ complexity.

System of equations (2.58) is solved separately for T different time periods, which is $\mathcal{O}(TL^3)$.

Consider now similar iterative algorithm for the series estimator. Note that the basis functions, exactly like kernel functions previously, have to be evaluated only once for each X_{tj} and can be kept in the memory in form of Ψ_t. Therefore this initial step is $\mathcal{O}(TKJ)$. In the following iterative steps one has to update $\alpha^{(r)}$ from (2.61), which is $\mathcal{O}(K^3(L+1)^3)$. Calculating new factors $Z_t^{(r)}$ is done separately for each t from (2.62), which leads to computational complexity of $\mathcal{O}(TL^3)$.

In the Newton-Raphson approach the argument of single calculation of Ψ_t is valid. The difference here lies in the fact that the update requires one step given by (2.63). The most crucial point involves calculating the inverse of $F'(\alpha^{OLD}, Z^{OLD})$. It means that each iteration step is $\mathcal{O}((L+1)^3(T+K)^3)$.

Comparing three estimation algorithms the initial part of evaluating the kernel or basis functions is linear in T, J, N and K. In particular when these functions have a complex form this part could be computationally intensive. Therefore it is worth to have a closer insight on the data design to exploit some possibly existing data patterns. For example, the design points could appear only in discrete place like in maturity direction in the implied volatility example, see Section 4.1.1.

In the iteration part all three methods are cubic in L. This, however should not be a great issue since L represents the dimension reduction level and has typically moderate size. The kernel method by separation of grid and time points is linear in N and T. It is not the case in series type of estimation where the estimation of basis functions have to be performed jointly. As a consequence this procedure is cubic in K. The iterative algorithm stays, however, linear in T, since the factors for each time point are estimated separately. In the Newton-Raphson case the update of (α, Z) is obtained simultaneously, so the algorithm is also cubic in T. This fact may lead to numerical difficulties in particular in case when $T \to \infty$.

2.9.2 Asymptotic Inference

In this section we present the asymptotic results essential for the inference on the estimated $\widehat{Z}_t$. We will show that, for any solution $\widehat{Z}_t$ and for any version of true Z_t, there exists a random matrix $\widehat{B}$ such that $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$ has asymptotically the same covariance structure as Z_t. This means that the difference of the inferences based on $\widetilde{Z}_t$ and Z_t is asymptotically negligible. This result is inline with (2.43), where one dimensional time series for parametric case is presented. Here we study the multivariate time series, which imposes additional complexity.

For the asymptotic analysis, we let K, J, $T \to \infty$. Our first result relies on the following assumptions.

(A1) The variables X_{11},...,X_{TJ}, ε_{11},...,ε_{TJ} are independent.

(A2) For $t = 1, ..., T$ the variables X_{t1},...,X_{tJ} are identically distributed, have support $\Xi = [0,1]^d$ and a density f_t that is bounded from below and above on $[0,1]^d$, uniformly over $t = 1, ..., T$.

(A3) We assume that

$$\begin{aligned} E[\varepsilon_{tj}] &= 0 \text{ for } t = 1, ..., T, j = 1, ..., J, \\ \sup_{t=1,...,T, j=1,...,J} E[\varepsilon_{tj}^2] &< \infty. \end{aligned}$$

(A4) The functions ψ_k may depend on the increasing indices T and J, but are normed so that $\int_{[0,1]^d} \psi_k^2(x)\, dx = 1$ for $k = 1, ..., K$.

(A5) The components $m_0, \ldots, m_L$ can be approximated by $\psi_1, \ldots, \psi_K$, i.e.

$$\delta_K = \sup_{x \in [0,1]^d} \inf_{A \in \mathbb{R}^{(L+1)\times K}} |m(x) - A\psi(x)| \to 0$$

as $K \to \infty$. We denote a matrix that fulfills $\sup_{x\in[0,1]^d} |m(x) - A\psi(x)| \leq 2\delta_K$ by A^*. We assume that $\delta_K = \mathcal{O}(K^{1/2} J^{-1/2})$ for $K, J \to \infty$.

(A6) There exist constants $0 < C_L < C_U < \infty$ such that all eigenvalues of the random matrix $T^{-1} \sum_{t=1}^T Z_t Z_t^\top$ lie in the interval $[C_L, C_U]$ with probability tending to one.

(A7) It holds that $(K \log K)/J \to 0$ and $\log T/J \to 0$.

(A8) The minimization (2.59) runs over all values of (A, z) with $z_{t,0} = 1$ and with

$$\sup_{x\in[0,1]} \max_{1\leq t\leq T} \|z_t^\top A\psi(x)\| \leq M_T,$$

where the constant M_T fulfils $\max_{1\leq t\leq T} \|Z_t\| \leq M_T/C_m$ for a constant C_m such that $\sup_{x\in[0,1]} \|m(x)\| < C_m$, and $M_T^2(K \log K/J) \to 0$, $M_T^2(\log T/J) \to 0$.

Condition (A8) and the additional bound M_T in the minimization is introduced for purely technical reasons.

Our first result gives rates of convergence for the least squares estimators $\widehat{Z}_t$ and $\widehat{A}$.

Theorem 2.9.3 *Suppose that model (2.47) holds and that $(\widehat{Z}_t, \widehat{A})$ is defined by the minimization problem (2.59) under the constraint (2.60). Make the assumptions (A1)–(A8). Then it holds that*

$$\frac{1}{T}\sum_{1\le t\le T}\left\|\widehat{Z}_t^\top\widehat{A} - Z_t^\top A^*\right\|^2 = \mathcal{O}_P(KJ^{-1}). \tag{2.67}$$

At this point we have made no assumptions on the sequence $Z_t : 1 \le t \le T$, besides the bound in (A8). Up to now it is allowed to be a deterministic or a random sequence. We now assume that it is a random process. We discuss how a statistical analysis differs if inference on Z_t is based on $\widehat{Z}_t$ instead of using (the unobserved) process Z_t. We will show that the differences are asymptotically negligible (except an orthogonal transformation). This is the content of the following theorem, where we consider estimators of autocovariances and show that these estimators differ only by second order terms. This asymptotic equivalence carries over to classical estimation and testing procedures in the framework of fitting a vector autoregresssive model. For the statement of the theorem we need the following assumptions:

(A9) The bound $\max_{1\le t\le T}\|Z_t\| \le M_T$ holds with probability tending to one, and it holds that $M_T^2\{(K\log K)/J\} \to 0$ and $M_T^2(\log T/J) \to 0$.

(A10) Z_t is strictly stationary with $E(Z_t) = 0$ and $E\|Z_t\|^\gamma < \infty$ for some $\gamma > 2$. It is strongly mixing with $\sum_{i=1}^\infty \alpha(i)^{(\gamma-2)/\gamma} < \infty$. The matrix $EZ_tZ_t^\top$ has full rank. The process Z_t is independent of $X_{11},\ldots,X_{TJ}$ and $\varepsilon_{11},\ldots,\varepsilon_{TJ}$.

(A11) The functions $m_0,\ldots,m_L$ are linearly independent. In particular, no function is equal to 0. Furthermore, it holds that $\sup_{x\in[0,1]}\|\psi(x)\| = \mathcal{O}(K^{1/2})$.

(A12) It holds that $K/J + \delta_K = o(T^{-1/2})$, $\log T = o(K)$, $K^5J^{-4}(\log K)^2 = o(T^{-1})$, and $K^7J^{-5}(\log K)^2 = o(T^{-1})$.

Condition (A10) implies that $T^{-1}\sum_{t=1}^T Z_t$ has a bounded second moment, see e.g. Corollary 1.1 in Bosq (1998).

Theorem 2.9.4 *Suppose that model (2.47) holds and that $(\widehat{Z}_t, \widehat{A})$ is defined by the minimization problem (2.59) under the constraint (2.60). Make the assumptions (A1)–(A12). Then there exists a random matrix $\widetilde{B}$, which is of the form (2.64), such that for $h \ge 0$*

$$\frac{1}{T}\sum_{t=h+1}^T\left(\widetilde{Z}_t - \overline{\widetilde{Z}}\right)\left(\widetilde{Z}_{t-h} - \overline{\widetilde{Z}}\right)^\top - \sum_{t=h+1}^T\left(Z_t - \overline{Z}\right)\left(Z_{t-h} - \overline{Z}\right)^\top = o_P(T^{-1/2}),$$

$$\overline{\widetilde{Z}} - \overline{Z} = o_P(T^{-1/2}),$$

where $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$, $\overline{\widetilde{Z}} = T^{-1}\sum_{t=1}^T \widetilde{Z}_t$ *and* $\overline{Z} = T^{-1}\sum_{t=1}^T Z_t$.

The proofs of Theorem 2.9.3 and Theorem 2.9.4 are given in Borak et al. (2007).

To illustrate an implication of Theorem 2.9.4, suppose that the factor process Z_t in (2.47) is a stationary VAR(p) process in a mean adjusted form

$$Z_t - \mu = \mathcal{A}(Z_{t-1} - \mu) + \cdots + \mathcal{A}_p(Z_{t-p} - \mu) + U_t, \tag{2.68}$$

where $\mu = E(Z_t)$, Θ_j is a $(L \times L)$ matrix of coefficients and U_t is a white noise with some nonsingular covariance matrix Σ_U. Here, we take the L components of Z_t, omitting the first $Z_{t,0} \equiv 1$. In a slight abuse of notation, we continue to refer to the resulting vector as Z_t.

Let Γ_h be a autocovariance matrix of the process Z_t with the lag $h = 0, 1, \ldots$, which can be estimated by

$$\widehat{\Gamma}_h = \frac{1}{T}\sum_{t=h+1}^{T} (Z_t - \overline{Z})(Z_{t-h} - \overline{Z})^\top.$$

For $Z_1, \ldots, Z_T$ rewrite (2.68) in a matrix notation as

$$\mathcal{Y} = \Theta\mathcal{Z} + U,$$

where $\mathcal{Y} = (Z_{p+1} - \mu, \ldots, Z_T - \mu)$, $\Theta = (\mathcal{A}_1, \ldots, \mathcal{A}_p)$ and $U = (U_{p+1}, \ldots, U_T)$. Define for $t = p, \ldots, T-1$ $\mathcal{Z}_t^0 = (Z_t - \mu, \ldots, Z_{t-p+1} - \mu)^\top$ then $\mathcal{Z} = (\mathcal{Z}_p^0, \ldots, \mathcal{Z}_{T-1}^0)$. From (2.16) the estimate of Θ is given by

$$\widehat{\Theta} = \widehat{\mathcal{Y}}\widehat{\mathcal{Z}}\left(\widehat{\mathcal{Z}}\widehat{\mathcal{Z}}^\top\right)^{-1},$$

where $\widehat{\mathcal{Y}}$ and $\widehat{\mathcal{Z}}$ are corresponding estimates of $\mathcal{Y}$ and $\mathcal{Z}$ with true mean μ replaced by a sample mean $\overline{Z}_t$. Likewise, fitting a VAR(p) model with estimated factors $\widetilde{Z}_t$ yields

$$\widetilde{\Theta} = \widetilde{\mathcal{Y}}\widetilde{\mathcal{Z}}\left(\widetilde{\mathcal{Z}}\widetilde{\mathcal{Z}}^\top\right)^{-1},$$

where $\widetilde{\mathcal{Y}}$ and $\widetilde{\mathcal{Z}}$ are defined as $\widehat{\mathcal{Y}}$ and $\widehat{\mathcal{Z}}$ with Z_t being replaced by $\widetilde{Z}_t$. Following (2.18) and (2.19) the estimators $\widehat{\Theta}$ and $\widetilde{\Theta}$ are composed of $\widehat{\Gamma}_h$ and $\widetilde{\Gamma}_h$ for different h, where

$$\widetilde{\Gamma}_h = \frac{1}{T}\sum_{t=h+1}^{T} (\widetilde{Z}_t - \overline{\widetilde{Z}})(\widetilde{Z}_{t-h} - \overline{\widetilde{Z}})^\top.$$

Theorem 2.9.4 states that $\sqrt{T}\left(\widetilde{\Gamma}_h - \widehat{\Gamma}_h\right) = o_P(1)$. Therefore $\sqrt{T}\left(\widetilde{\Theta} - \Theta\right) = o_P(1)$. This together with (2.17) gives

$$\begin{aligned}\sqrt{T}\left(\widetilde{\Theta} - \Theta\right) &= \sqrt{T}\left(\widetilde{\Theta} - \widehat{\Theta}\right) + \sqrt{T}\left(\widehat{\Theta} - \Theta\right) \\ &= o_P(1) + \mathcal{O}_P(1) = \mathcal{O}_P(1).\end{aligned} \tag{2.69}$$

From (2.69) the error of the inference based on the $\widetilde{Z}_t$ can be decomposed to the error of the DSFM estimation and error of the inference based on the unobservable Z_t. Asymptotically, however, the first error is negligible in compare to the second one. Therefore the asymptotic inference based on $\widetilde{Z}_t$ is equivalent to the inference based on Z_t.

2.9.3 Model Selection

Similarly to the case of the traditional factor model, see Section 2.4, there are some identification issues that have to be addressed. As we have pointed out above for the pair of estimates $(\widehat{Z}_t, \widehat{m})$ the corresponding pair $(\widetilde{B}^\top \widehat{Z}_t, \widetilde{B}^{-1}\widehat{m})$ yields the same fit, where the form of $\widetilde{B}$ is in (2.64). In particular this issue underline the impossible sign detection of the factors and corresponding loadings. Another identification issue is shifting an intercept function by scaled function m_l and subtracting a constant from the time series of factors

$$\begin{aligned}m_0(x) + \sum_{l=1}^{L} Z_{tl} m_l(x) &= m_0(x) + \sum_{l=1}^{L} \{(Z_{tl} - c_l) m_l(x) + c_l m_l(x)\} \\ &= m_0(x) + \sum_{l=1}^{L} c_l m_l(x) + \sum_{l=1}^{L} Z_{tl}^{new} m_l(x) \\ &= m_0(x)^{new} + \sum_{l=1}^{L} Z_{tl}^{new} m_l(x).\end{aligned}$$

One possible setting for the constants c_l is such that $\widehat{Z}_t$ has a zero mean. For the simplicity of notation Z_t is understood without the first unit component $Z_{t0} = 1$.

For the model identification one can always choose $\widehat{m}$ such that $\widehat{m}_1, \ldots, \widehat{m}_L$ are orthogonal in $L_2([0,1]^d)$ or in other L_2, e.g. in $L_2(T^{-1}\sum_{t=1}^{T} \widehat{p}_t)$ where $\widehat{p}_t$ is a kernel estimate of the density of X_{tj}. Another possibility is to project on the obtained $\widehat{m}_l$ on specified functions m_l, which comes from external practical motivation. One may try to be close to some parametric form of the functions, which may yield convenient interpretation, e.g. level, slope and

curvature. This leads to minimizing in the appropriate space $\| m - \widetilde{B}^{-1}\widehat{m} \|$. In this case matrix B^{-1} has the form

$$B^{-1} = \mathbf{M}\widehat{\mathbf{M}}^{-1},$$

where $\mathbf{M}$ and $\widehat{\mathbf{M}}$ are defined by its elements $\mathbf{M} = (\int m_{l'}\widehat{m}_{l''})_{l',l''}$ and $\widehat{\mathbf{M}} = (\int \widehat{m}_{l'}\widehat{m}_{l''})_{l',l''}$ for $l', l'' = 1, \ldots, L$. In Theorem 2.9.4 we state that there exists matrix $\widetilde{B}$, for which the covariance structure of a $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$ converge to the covariance structure of unobservable Z_t. We choose $\widetilde{B}$ as

$$B = \left(T^{-1} \sum_{t=1}^{T} Z_t \widehat{Z}_t^\top \right)^{-1} T^{-1} \sum_{t=1}^{T} Z_t Z_t^\top. \tag{2.70}$$

The form of the matrix is motivated by the technical proof of the Theorem 2.9.4 in Borak et al. (2007). This selection may also serve for the identification purpose. We use this idea in simulation study of Section 3.2, where the true data generating process is known.

Since the model is not nested, the number of the dynamic functions needs to be determined in advance. One of the possible solution is to consider

$$RV(L) = \frac{\sum_t^T \sum_j^{J_t} \{Y_{t,j} - \sum_{l=0}^{L} \widehat{Z}_{t,l}\widehat{m}_l(X_{t,j})\}^2}{\sum_t^T \sum_j^{J_t} (Y_{t,j} - \bar{Y})^2}. \tag{2.71}$$

The quantity $1 - RV(L)$ can be interpreted as a proportion of the variation explained by the model among the total variation and belongs to the goodness of fit measures. One may also construct AIC or BIC type of criterion, where one penalizes the number of the dynamic functions in the model. Another possible idea is to fit the model on subsamples. In this framework one may perform some type of cross-validation. The full cross-validation seems to be infeasible for the large samples but removing groups of data, e.g. all observation with common time index t, can be possibly less computationally intensive. The alternative approach is to divide the sample into training an evaluation parts and minimize empirical prediction error with respect to L. This idea, however, requires a statistical model for the process $\widehat{Z}_t$.

For the data driven choice of bandwidth in the kernel method we refer to Fengler et al. (2007). They propose to minimize with respect to bandwidths

$$\Xi_{AIC_1} = \sum_{t,j} \left\{ Y_{tj} - \sum_{l=0}^{L} \widehat{Z}_{tl}\widehat{m}_l(X_{tj}) \right\}^2 w(X_{tj}) \exp\left\{ \frac{2L}{N} \mathcal{K}_h(0) \int w(u)du \right\}$$

or computationally more easy criterion

$$\Xi_{AIC_2} = \sum_{t,j} \left\{ Y_{tj} - \sum_{l=0}^{L} \widehat{Z}_{tl}\widehat{m}_l(X_{tj}) \right\}^2 \exp\left\{ \frac{2L}{N} \mathcal{K}_h(0) \frac{\int w(u)du}{\int w(u)\widehat{p}(u)du} \right\},$$

where $\widehat{p}(u) = T^{-1} \sum_t \widehat{p}_t(u)$. One possible selection of the weight function is $w(u) = \frac{1}{\widehat{p}(u)}$.

The series estimation method is in this thesis reduced to B-spline estimators. The estimates depend on spline order, knots number and their location. The traditional ways for placing the knots are equidistant knots and knots equally spaced in sample quantiles, see Ruppert et al. (2003). In order to avoid placing the knots in the empty regions of the domain, in the DSFM estimation one has to additionally take into account time development. This could be achieved by placing equidistant knots in transformed space, with transformation based on the marginal empirical distribution functions of X_{tj} for all t and j. The spline order and number of functions K can be optimized by using criteria discussed above. One may also build forward selection, backward elimination, or stepwise procedures based on introducing or deleting the knots from the knots candidates.

Chapter 3

Sensitivity Analysis

"Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin."

John von Neumann

The previous chapter focusses on the theoretical background of the DSFM. It presents how the model is related to other statistical techniques, show estimation methodology, asymptotic results and some theoretical properties. This chapter studies the model from the practical perspective and presents small sample properties. The estimates are obtained on the simulated data in order to perform the sensitivity analysis with respect to changes of an input setting.

The complexity of many statistical problems hampers very often the study of their properties. This difficulty may be overcome with simulations. Although they yield only an approximative answer this approximation suffice in many cases. Due to increasing computational power more complex computer experiments have become feasible. Nevertheless one will ever face the problem with tradeoff of accuracy and speed.

In this chapter we present the simulation based studies on the properties of the DSFM. Due to the nonparametric structure and the iterative estimation procedure these properties are hard to infer analytically. We believe that the performed simulations conveniently reflect the true behavior of the model. Our aim here is twofold. First, the simulations illustrate some results discussed already in Chapter 2. They provide a closer insight and serve as an additional explanation tool. Second, we study the estimates of the DSFM on different data structure in order to infer some conclusions on its behavior.

The profusion for the DSFM simulations' settings is almost unlimited. One may choose different Z_t, m, design for X_t etc. Here, we focus only on some particular cases and believe that they are vital and interesting. We try to have a nontrivial structure, which remains perceptible. Nevertheless there is always a room for more experiments so the presented results can be never fully generalized.

In Section 3.1 we introduce the simulation setting. Section 3.2 illustrates the asymptotic issues presented in Theorem 2.9.4. The simulation's parameters sensitivity is studied in Section 3.3 and forecasting experiment is performed in Section 3.4. In Section 3.5 the dependence on the data design is considered.

3.1 Simulation Settings

Since the main aim of this chapter is to investigate the DSFM on the basis of simulation experiment it is vital to give an overview on designing the sampling method. This section presents only general idea of constructing the simulations for the DSFM. However, this pattern is reproduced in all following sections with different settings and constraints.

The main idea of each simulation experiment conveys the usage of random data. The data generating process, however, is fully known and one can indicate the all sources of randomness. Then the result of the simulation is treated as an observed sample and the estimation procedure is applied to these artificial data. The full knowledge on the true model allows to compare the estimates to the truth and draw conclusions on the model's performance.

For the more formal presentation assume a statistical model $\mathcal{P}_\theta$ and generate I samples of size n $(X_1, \ldots, X_n)^{(i)}$ from $\mathcal{P}_\theta$ for $i = 1, \ldots, I$. In the next step estimate $\widehat{\theta}^{(i)}$ and compare them to θ. One possible way is to calculate values of some function $g(\widehat{\theta}^{(1)}, \ldots, \widehat{\theta}^{(I)}, \theta)$. The function g has to be intuitive and meaningful in order to get some insight. For this purpose often the graphical tools can be used to get clear representation of the results. In particular it could be worth to present g as a function of θ or n. In a similar way some sensitivity to the estimation procedure parameters, e.g. bandwidths, could be studied, since they influence $\widehat{\theta}^{(i)}$. In this chapter we often rely on the comprehensive plots rather than pure numbers since we believe that they bring valuable intuition. For the importance and efficient usage on graphical techniques we refer to Chen et al. (2007).

The setting for the DSFM follows the presented scheme. In order to get the random sample (X_{tj}, Y_{tj}) for $j = 1, \ldots, J_t$ and $t = 1, \ldots, T$ one has to set the design points X_{tj} from some selected set Ξ. They can follow the random

design with the density f_t or fixed design. Next the points Y_{tj} have to be simulated according to some model. In order to fix the model one has to specify $L+1$ functions operating on Ξ and some process for Z_t to generate the dynamics. At the end Y_{tj} is given from (2.47), where some additional error ε_{tj} is also included.

To be more specific we present the simulation setting used later in Section 3.2 - Section 3.5. For the points X_{tj} we take some distribution (in most cases uniform) on the unit square $\Xi = [0,1]^2$ and draw them independently. We consider the following tuple of 2-dimensional functions

$$\begin{pmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{pmatrix}(x_1, x_2) = \begin{pmatrix} 1 \\ 3.46(x_1 - \frac{1}{2}) \\ 9.45\left\{(x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2\right\} - 1.6 \\ 1.41\sin(2\pi x_2) \end{pmatrix}. \tag{3.1}$$

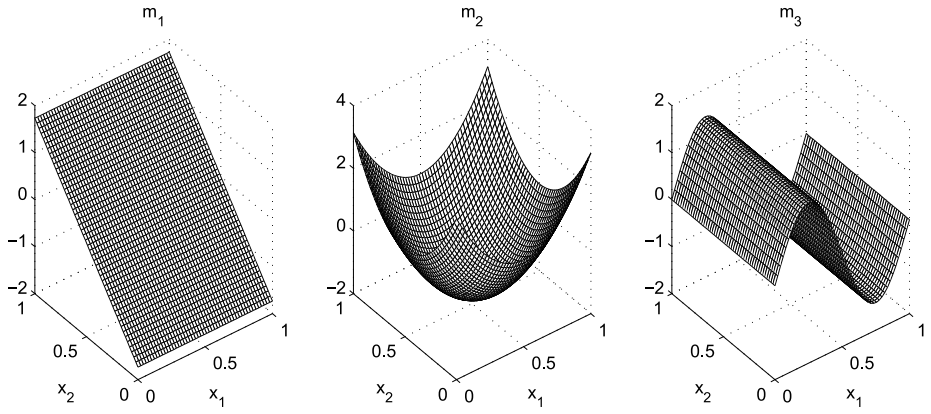

Figure 3.1: *True functions m_1, m_2, m_3 from which the data are generated in simulation examples.*

The coefficients in (3.1) are chosen so that m_1, m_2, m_3 are close to orthogonal. The functions are displayed in Figure 3.1. The time series Z_t follows VAR(1) process $Z_t = \mathcal{A}Z_{t-1} + U_t$, where U_t is $N_3(0, \Sigma_U)$ random vector and

$$\mathcal{A} = \begin{pmatrix} 0.95 & -0.2 & 0 \\ 0 & 0.8 & 0.1 \\ 0.1 & 0 & 0.6 \end{pmatrix}, \quad \Sigma_U = \begin{pmatrix} 10^{-4} & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 10^{-4} \end{pmatrix}.$$

The functions in (3.1) are chosen such that the generated surfaces reflect 'interesting' structure. Two typical surface are presented in Figure 3.2. One can clearly recognize a strong impact of the first and third factor in the left panel and bigger influence of the second factor in the right panel.

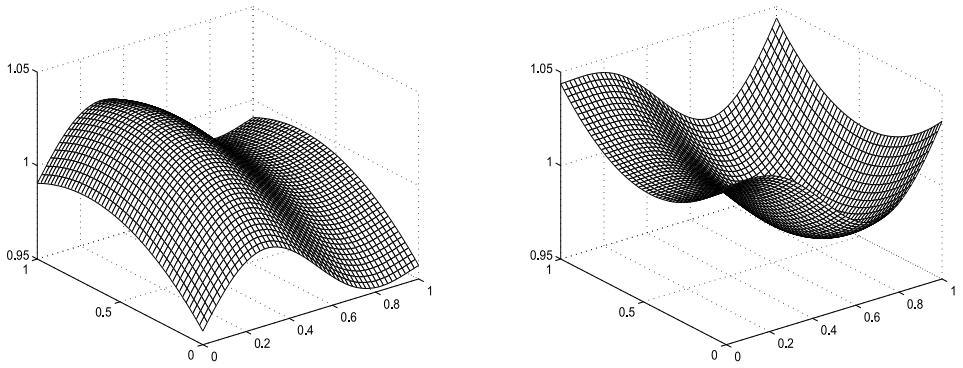

Figure 3.2: *Examples of surfaces that could be generated in simulation experiments.*

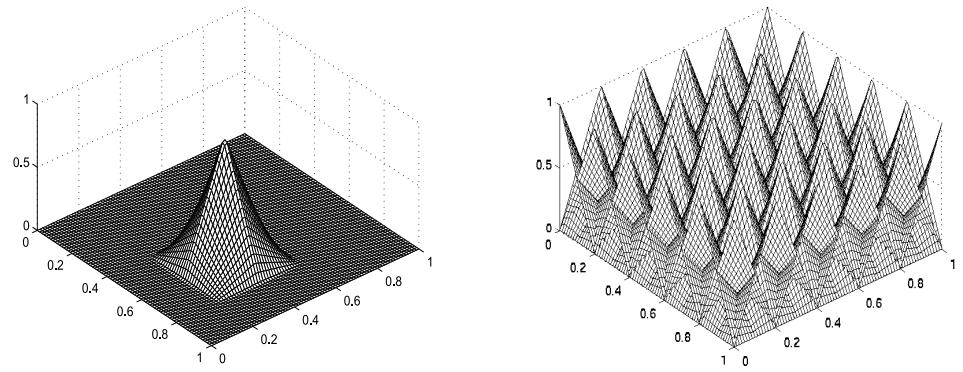

Figure 3.3: *Tensor linear B-spline basis used in the estimation. Left panel: one particular basis function ψ_k. Right panel: the whole set of basis functions for $K = 36$.*

In order to obtain the DSFM estimates one has to apply some estimation method described in Section 2.9.1. For the numerical convenience we choose series estimator with iterative updates for m and Z_t. For the estimation we employ, the tensor products of B-splines. For example, the one-dimensional linear B-splines $\widetilde{\psi}_k$ are defined on a consecutive equidistant knots x^k, x^{k+1}, x^{k+2} by $\widetilde{\psi}_k(x) = (x - x^k)/(x^{k+1} - x^k)$ for $x \in (x^k, x^{k+1}]$, $\widetilde{\psi}_k(x) = (x^{k+2} - x)/(x^{k+2} - x^{k+1})$ for $x \in (x^{k+1}, x^{k+2}]$, and $\widetilde{\psi}_k(x) = 0$ otherwise, see also Section 2.1. The tensor spline basis functions in case $K = 36$ are plotted in Figure 3.3. In the simulation we increased the number of the basis functions K to 49 and 64, which correspond to more dense layouts in the right panel of Figure 3.3.

3.2 Asymptotic Inference

In Section 2.9.2 we argue that the inference based on the covariances of the unobserved factors Z_t is asymptotically equivalent to the one based on $\widetilde{B}^\top \widehat{Z}_t$ for some invertible $\widetilde{B}$. The form of $\widetilde{B}$ is given in (2.70). In this section we illustrates the equivalence by a simulation study. For this, we compare the covariances of Z_t and $\widetilde{Z}_t \equiv \widetilde{B}^\top \widehat{Z}_t$.

We take $T = 500, 1000, 2000$, $J = 100, 250, 1000$ and $K = 36, 49, 64$ of linear tensor B-splines. We consider the simulation setting described in Section 3.1. For ε_{tj} are i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.05$. Each simulation is repeated 250 times for each combination of (T, J, K).

We plot in Figure 3.4 the entries of the scaled difference of the covariance matrices

$$\widetilde{D} = \frac{1}{\sqrt{T}} \left\{ \sum_{t=1}^{T} \left(\widetilde{Z}_t - \overline{\widetilde{Z}} \right) \left(\widetilde{Z}_t - \overline{\widetilde{Z}} \right)^\top - \sum_{t=1}^{T} \left(Z_t - \overline{Z} \right) \left(Z_t - \overline{Z} \right)^\top \right\}. \quad (3.2)$$

Each panel of Figure 3.4 corresponds to one entry of the matrix $\widetilde{D}$, and the three boxplots in each panel represent the distributions of the 250 values of the corresponding entry for $T = 500, 1000, 2000$. In the figure we also depicted, by thick lines, the 95% and 5% quantiles of

$$D = \frac{1}{\sqrt{T}} \left\{ \sum_{t=1}^{T} \left(Z_t - \overline{Z} \right) \left(Z_t - \overline{Z} \right)^\top - T\Gamma \right\}, \quad (3.3)$$

where Γ is the true covariance matrix of the simulated VAR process. It is known, see Section 2.5.1, that Γ can be represented as $\gamma = (I_{L^2} - \mathcal{A} \otimes \mathcal{A})^{-1} \sigma_U$, where γ and σ_U are the stack forms of Γ and Σ_U respectively.

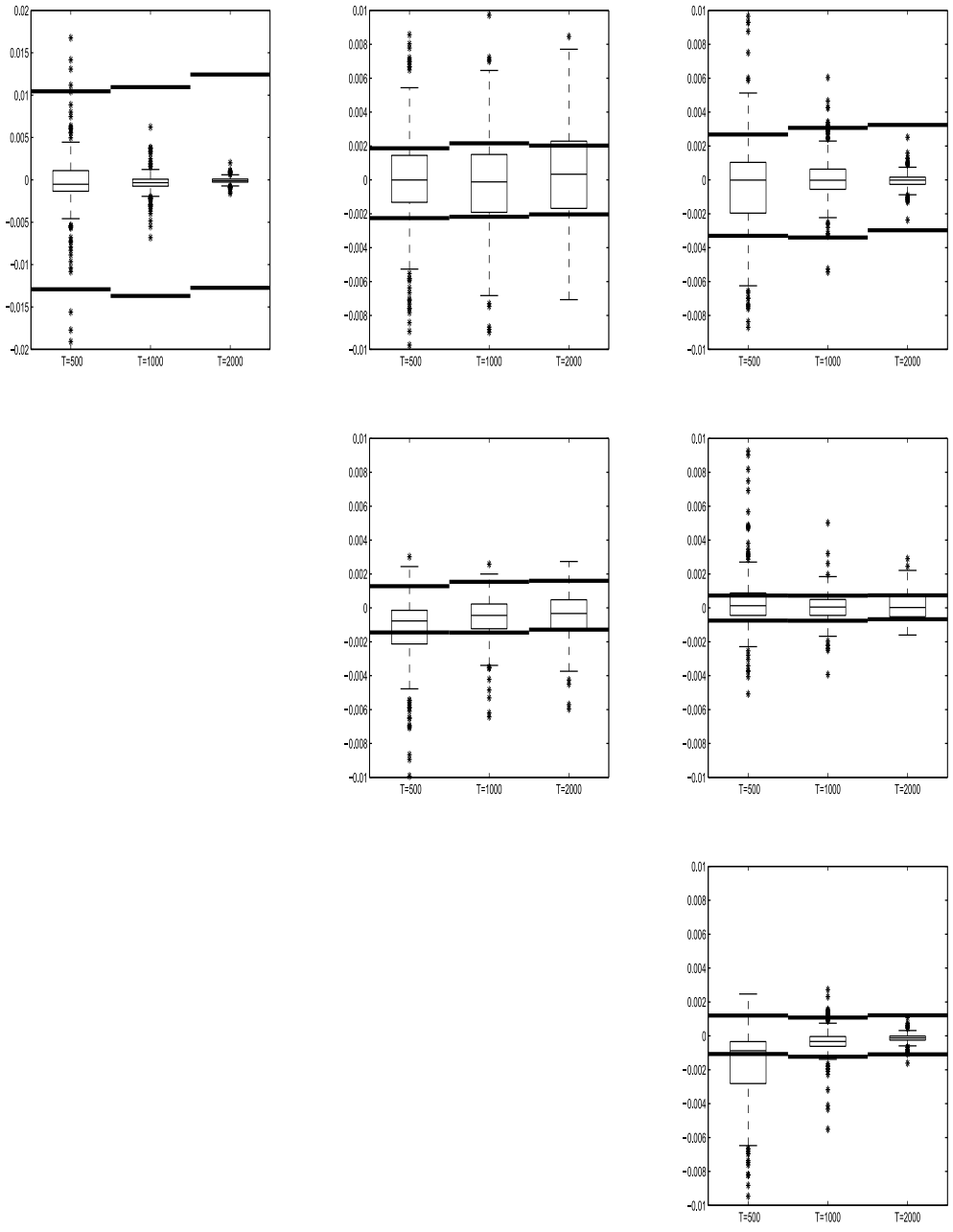

Figure 3.4: *The boxplots based on 250 values of the entries of the scaled difference of the covariance matrices given at (3.2). The lengths of the series* Z_t *and* $\widetilde{Z}_t$ *were* $500, 1000, 2000$. *The thick lines represent the* 95% *and* 5% *quantiles of (3.3).*

The theory in Section 2.9.2 tells that the size of $\widetilde{D}$ is of smaller order than the normalized error D of the covariance estimator based on Z_t. It is known that the latter converges to a non-degenerate law as $T \to \infty$. This is well supported by the plots in Figure 3.4 showing that the distance between the two thick lines in each panel is invariant as T increases. The fact that the additional error incurred by using $\widetilde{Z}_t$ instead of Z_t is negligible for large T is also confirmed. In particular, the long stretches at tails of the distributions of $\widetilde{D}$ as well as their interquartile ranges get shorter as T increases.

3.3 Parameters Sensitivity

In this section we check the sensitivity of the DSFM estimates with respect to different parameters. After fixing the model, as described in Section 3.1, we generate samples with varying variance σ^2, number of points per time period J and the time series length T. Additionally, we check the performance of the method when the number of splines K and their order is changing.

For the evaluation of the results two types of errors are defined for each simulation i and time point t. First we consider the mean square error (MSE) given as

$$MSE(t,i) = \frac{1}{J}\sum_{j=1}^{J}\left\{Y_{tj}^{(i)} - Z_t^{(i)\top} m^{(i)}(X_{tj}^{(i)})\right\}^2,$$

where $Y_{tj}^{(i)}$, $X_{tj}^{(i)}$, $Z_t^{(i)}$ $m^{(i)}$ correspond to the ith repetition in time point t. Similarly one may define $MSE(t,\cdot) = \frac{1}{I}\sum_i MSE(t,i)$ as an error at point t and overall MSE as $MSE = \frac{1}{T}MSE(t,\cdot)$. MSE checks how well the model could replicate the variables Y_{tj}. However, in the simulation experiment the true model from which the data are generated is fully known and therefore one may in similar manner define the error as a discrepancy from this true model. One possible approach conveys the consideration of the mean integrated square error (MISE) defined as

$$MISE(t,i) = \int_{\Xi}\left(Z_t^{(i)\top} m(u) - \widetilde{Z}_t^{(i)\top}\widetilde{m}^{(i)}(u)\right)^2 du, \tag{3.4}$$

where $Z_t^{(i)}$ and m reflect the data generating model and $\widetilde{Z}_t^{(i)}$, $\widetilde{m}^{(i)}$ some estimates. $MISE(t,\cdot)$ and $MISE$ can be defined accordingly. Obviously for the practical calculation of the integral in (3.4) some discrete scheme has to be applied.

In this section we consider three types of the estimates. By referring to $(\widehat{Z}_t, \widehat{m})$ we mean the estimates obtained through the DSFM. $\check{Z}_t$ describe the

case where only the factors are estimated and one exploits the knowledge of the true functions m. This estimate is simply given by (2.62) where the true functions replace the corresponding series estimator. Accordingly, $\check{m}$ denotes the estimates with Z_t given at hand, obtained by direct application of (2.61).

3.3.1 Noise

The modelling challenges contain the proper recognition of signal-to-noise ratio. Here we study the performance of the model with varying intensity of the noise. We set ε_{tj} i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.005$, 0.01, 0.05, 0.1, 1, 2. For each level of variance we perform 100 simulations with $T = 1000$, $J = 250$ and $K = 49$ of quadratic tensor B-splines.

Figure 3.5 presents $MISE(t, \cdot)$ for three different types of estimation. The left panel shows the results obtained by applying the DSFM. The middle panel corresponds to the case with true functions m and $\check{Z}_t$ plugged in (3.4) for $\widetilde{m}$ and $\widetilde{Z}_t$ respectively, while the right one presents the estimation of the functions only. Note that the right panel have different scale than other figures. All three plots confirm that while the level of noise increases the recovery of the correct signal is becoming more demanding. It can be seen by the location of the $MISE(t, \cdot)$. It is also visible in Figure 3.6 where $MISE$ is plotted on the double logarithmic scale. In the both figures one sees that $MISE$ is smaller for $\check{Z}_t$ and $\check{m}$ than DSFM. In all cases the error is linear in variance of the noise. This fact can be easily checked by fitting a parabolic function but for the clarity of the plot we avoid this unnecessary complication.

σ	DSFM	true model	true m	true Z_t
0.005	$2.473 \cdot 10^{-5}$	$2.500 \cdot 10^{-5}$	$2.470 \cdot 10^{-5}$	$2.503 \cdot 10^{-5}$
0.01	$9.877 \cdot 10^{-5}$	$9.999 \cdot 10^{-5}$	$9.879 \cdot 10^{-5}$	$9.996 \cdot 10^{-5}$
0.05	$2.468 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.497 \cdot 10^{-3}$
0.1	$9.888 \cdot 10^{-3}$	$1.000 \cdot 10^{-2}$	$9.880 \cdot 10^{-3}$	$9.992 \cdot 10^{-3}$
1.0	$9.846 \cdot 10^{-1}$	1.000	$9.879 \cdot 10^{-1}$	$9.992 \cdot 10^{-1}$
2.5	6.145	6.248	6.179	6.244

Table 3.1: *MSE in a simulated example presented as a function of time for different level of noise.*

In addition to $MISE$ we report in Table 3.1 MSE, which reflects how well the model replicates the data. For this purpose we consider also the true model, i.e. both functions m and time series Z_t are known. The table shows that considering the true model and calculating MSE is in fact equivalent

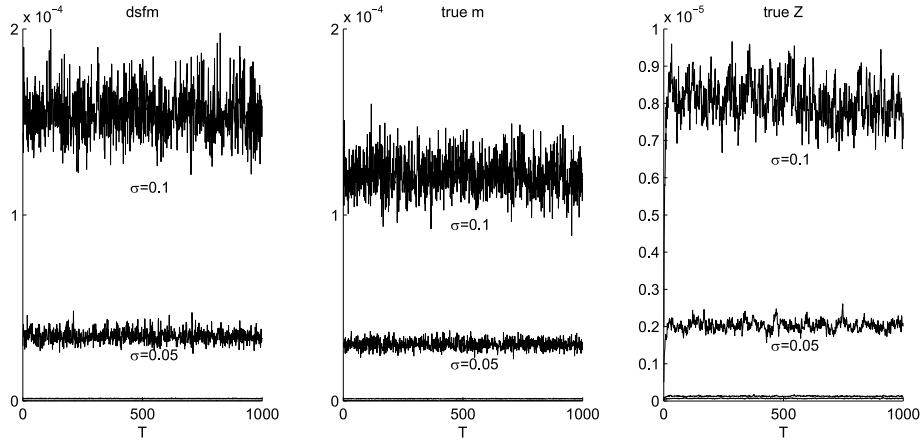

Figure 3.5: *MISE in a simulated example presented as a function of time for three different methods and different level of noise σ. Left panel: the DSFM. Middle panel: true functions m are known. Right panel: true time series Z_t are known. The plots indicate only $\sigma = 0.005$ and 0.1.*

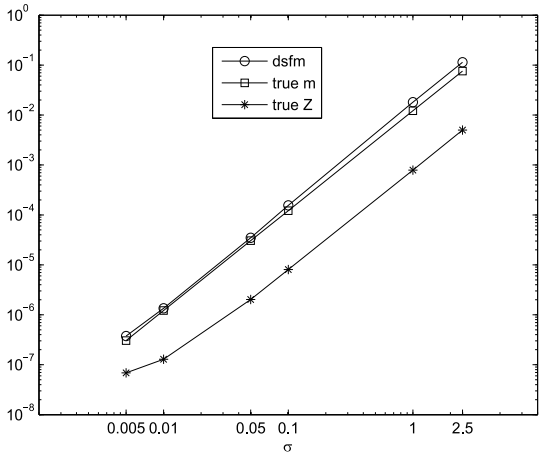

Figure 3.6: *MISE in a simulated example for different levels of noise and different methods presented in the double logarithmic scale.*

to estimating variance of the noise across all points for all simulations. This estimates are in fact quite close the true variance. Note that MSE for the true model is highest in compare to the cases where either Z_t, m or both are estimated. This correspond to the fact that the estimation procedures fits also partially some structure to the noise itself. The errors of the DSFM and $\check{Z}$ estimates are comparable and smaller to the $\check{m}$ estimates. This fact can be explained by the ratio parameters-to-observations. In estimating $\check{m}$ one has $L+1$ loading functions represented by K basis functions, which have to estimated from TJ observation. Estimation of $\check{Z}_t$ however is done for each time point separately i.e. L parameters are obtained from J observation. Usually $L/J > (L+1)K/TJ$ since K/T is of the moderate size. In the DSFM one posses more degrees of freedom since both m and Z_t need to be estimated. Therefore in the most cases it reaches smallest MSE.

3.3.2 Number of Observations

In the statistical modelling the larger sample size brings usually more information. In this section we would like to illustrate this fact by simulation study with increasing number of observations per time point J. Similarly to the previous section we compare the DSFM, $\check{m}$ and $\check{Z}_t$. We simulate according to the setting mentioned before with $\sigma = 0.05$ and $J = 100, 250, 500, 1000$. For the basis functions we take $K = 49$ quadratic tensor B-splines.

Figure 3.7 shows the behavior of $MISE$ with varying J. As the number of points increases the errors are decreasing and resemble hyperbolic decrease. The intuition for this fact comes from the ratio parameters-to-observations discussed in Section 3.3.1. The smallest error is for $\check{m}$. The errors for the DSFM and $\check{Z}_t$ have comparable magnitude, but estimating only $\check{Z}_t$ comes closer to the true model since the functions are already correctly specified.

The descriptive statistics of MSE presented in Table 3.2 show that the fit of the data done with true model is invariant to J. For other cases the error is increasing with the number of points. It cannot be surprising because with larger sample size and fixed number of parameters, the estimates converge to the true model, in the spanned space sense. With the moderate sample size less smoothing is involved, which results in closer fit to the data.

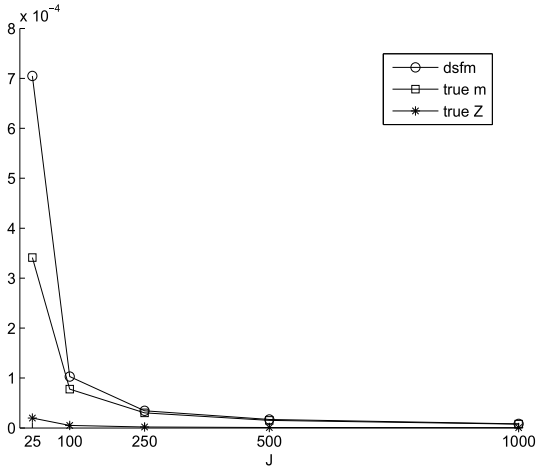

Figure 3.7: *MISE in a simulated example for different number of observations in one particular time point.*

J	DSFM	true model	true m	true Z_t
25	$2.312 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$	$2.200 \cdot 10^{-3}$	$2.479 \cdot 10^{-3}$
100	$2.437 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.425 \cdot 10^{-3}$	$2.495 \cdot 10^{-3}$
250	$2.469 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.497 \cdot 10^{-3}$
500	$2.484 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.485 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$
1000	$2.492 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$	$2.492 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$

Table 3.2: *MSE in a simulated example for different number of observations in one particular time point.*

3.3.3 Time Series Length

In Section 3.3.2 we discuss the case when the sample size is increasing by the changing the number of observations per time point. Here we keep fixed $J = 250$ and increase the length of the factor time series to $T = 250, 500, 1000, 2000$. The other simulation settings are same to Sections 3.3.2.

Figure 3.8 presents $MISE$. One may observe the similar behavior as in Figure 3.3.2, except $\check{Z}_t$ case. The precision of the estimate does not depend on T so the error is constant in T. This fact can be also illustrated by the behavior of MSE, see Table 3.3. For the true m MSE is stable like in the case of the true model. It varies only when some part of the estimation is performed on the pooled data.

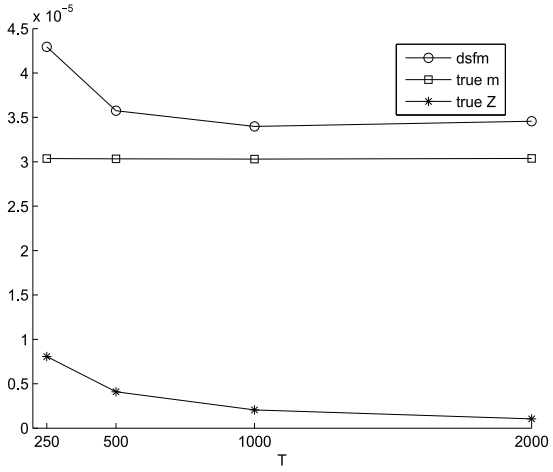

Figure 3.8: *$MISE$ in a simulated example for different number of time points.*

T	DSFM	true model	true m	true Z_t
250	$2.466 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.470 \cdot 10^{-3}$	$2.492 \cdot 10^{-3}$
500	$2.468 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.495 \cdot 10^{-3}$
1000	$2.469 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.470 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
2000	$2.470 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.470 \cdot 10^{-3}$	$2.499 \cdot 10^{-3}$

Table 3.3: *MSE in a simulated example for different number of time points.*

3.3.4 Splines

In the previous sections we study the sensitivity with respect to the different data structure by keeping the estimation algorithm fixed. In this section we fix the data setting and pay attention to the performance of the DSFM with different choice of the basis functions. We take $K = 25, 36, 49, 64$ tensor B-splines and consider as well different order - constant, linear, quadratic and cubic. The data are simulated according to the setting presented in Section 3.1 with $J = 250$, $T = 1000$ and $\sigma = 0.05$.

Figure 3.9 shows $MISE$ and Table 3.4 presents MSE. Note that in cases when the functions are not estimated the error have to be invariant of the spline choice, see $MISE$ and MSE of the true m and MSE of the true model. For the cases where the m functions' estimation is relevant, only for constant basis functions their number have significant meaning. Figure 3.9 indicates that while for the constant splines the error drops drastically with the increase of K, it stays relatively stable for linear, quadratic and cubic splines. Moreover, the differences in error are negligible with respect to the order of splines and it is not evident that higher order use to yield better fit.

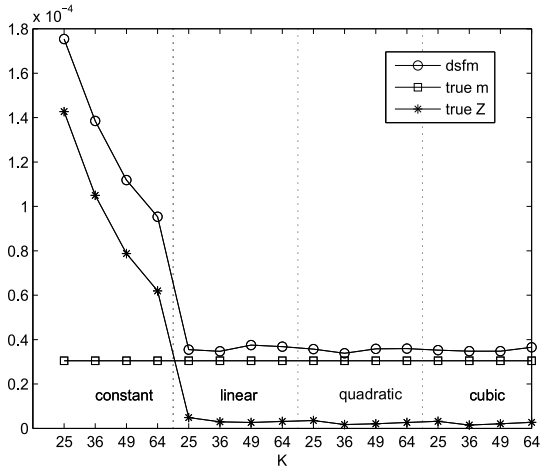

Figure 3.9: *$MISE$ in a simulated example for various setting of tensor B-splines. The dotted vertical lines separates the order of splines. The error is presented as a function of the number of basis functions K.*

order	num. splines	DSFM	true model	true m	true Z_t
1	25	$2.611 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.642 \cdot 10^{-3}$
1	36	$2.568 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.598 \cdot 10^{-3}$
1	49	$2.542 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.572 \cdot 10^{-3}$
1	64	$2.524 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.553 \cdot 10^{-3}$
2	25	$2.472 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.501 \cdot 10^{-3}$
2	36	$2.471 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
2	49	$2.470 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
2	64	$2.470 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.497 \cdot 10^{-3}$
3	25	$2.473 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.501 \cdot 10^{-3}$
3	36	$2.470 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
3	49	$2.471 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
3	64	$2.469 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.497 \cdot 10^{-3}$
4	25	$2.472 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.450 \cdot 10^{-3}$
4	36	$2.471 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
4	49	$2.470 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.498 \cdot 10^{-3}$
4	64	$2.470 \cdot 10^{-3}$	$2.500 \cdot 10^{-3}$	$2.469 \cdot 10^{-3}$	$2.497 \cdot 10^{-3}$

Table 3.4: *MSE in a simulated example for various setting of tensor B-splines. Different spline's order and number of basis functions is considered.*

3.4 Forecasting Experiments

In Chapter 2 and particularly in Theorem 2.9.4 we argue that the asymptotic inference based on the true time series Z_t is equivalent to the inference based on the estimated time series $\widehat{Z}_t$. There exist a random matrix $\widetilde{B}$ such that covariance structure of $\widetilde{Z}_t = \widetilde{B}^\top \widehat{Z}_t$ converge to the covariance structure of Z_t. This fact has been already illustrated in Section 3.2. Here, we show a practical meaning of the mentioned theorem. Our illustration covers the forecasting experiment, where we compare the out-of-sample error based on the true time series and estimated time series. Note that this kind of study is only possible in the simulation framework, where the true model is fully specified.

For our experiment we generate the data almost exactly like in Section 3.2, i.e. we take $T = 250, 1000, 2000$ and $J = 100, 250, 1000$. Based on the results in Section 3.3.4 we keep $K = 49$ of quadratic tensor B-splines. Each simulation setting is repeated 100 times.

In contrast to the previous sections the data are generated for longer time horizon up to time $T + H$, where $H = 150$. The estimation algorithm is employed, however, only up to time T. The rest of the sample serves for pre-

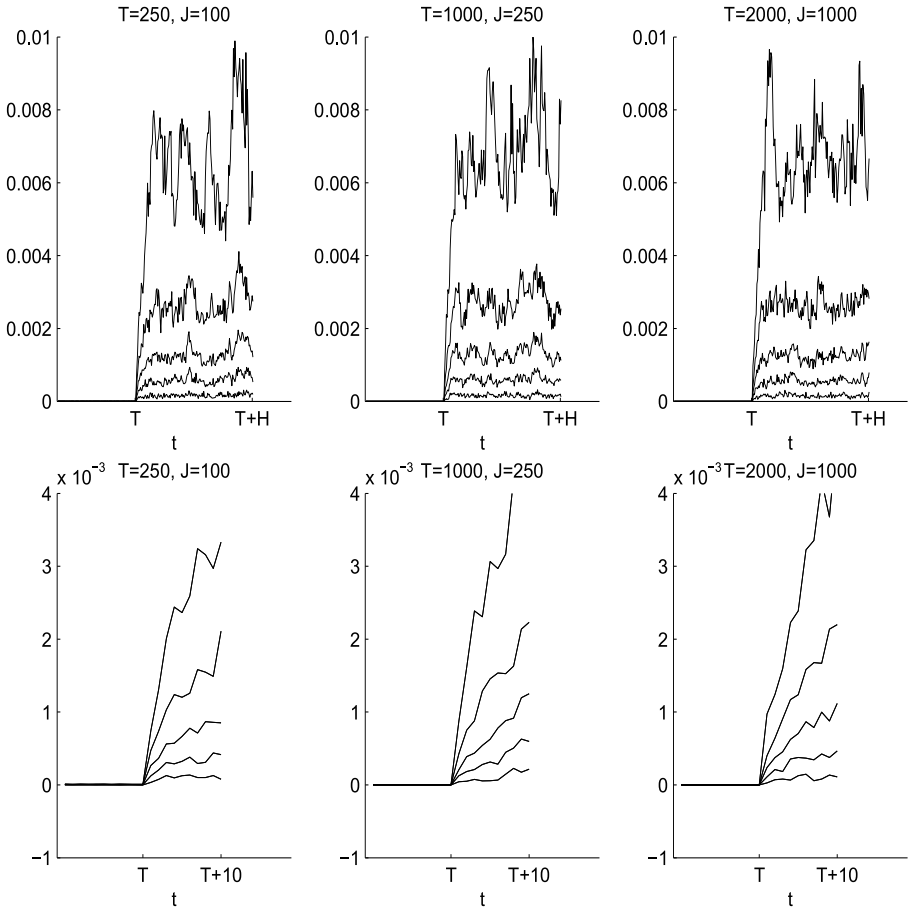

Figure 3.10: *Quantiles* $(5\%, 25\%, 50\%, 75\%, 95\%)$ *of in-sample and out-of-sample errors in a simulated example. Upper panel present the quantiles up to* $H = 150$*. Lower panel are the magnification for short time periods. Different settings of* T *and* J *are given:* $(T = 250, J = 100)$, $(T = 1000, J = 250)$, $(T = 2000, J = 1000)$ *looking rightwards.*

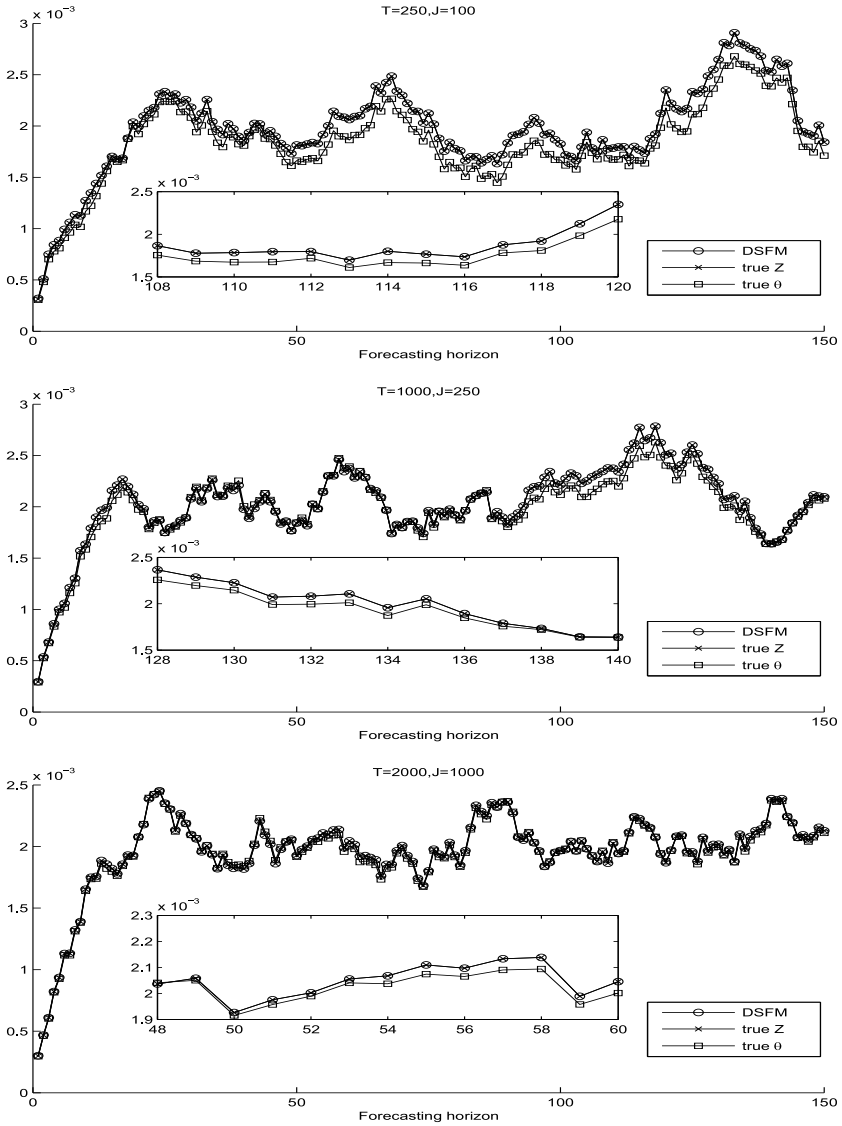

Figure 3.11: *Comparison of forecasting errors for different methods. The DSFM forecast is confronted with the true time series Z_t and true parameter θ. Different settings of T and J are given: $(T = 250, J = 100)$, $(T = 1000, J = 250)$, $(T = 2000, J = 1000)$ looking downwards.*

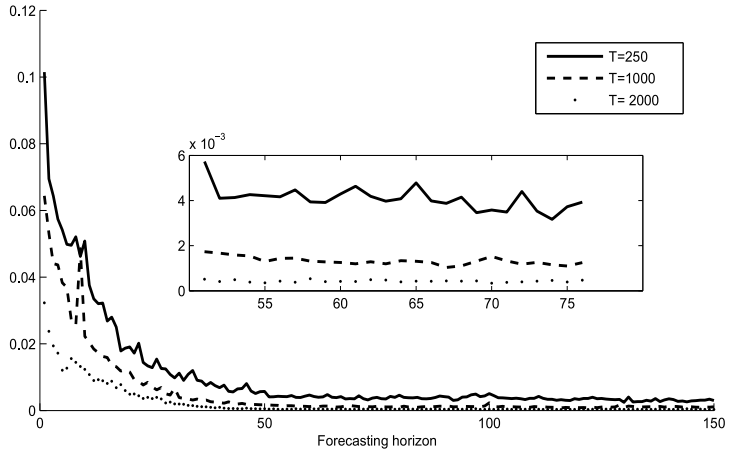

Figure 3.12: *Absolute relative differences of forecasting errors based on the DFSM and true time series. The results are displayed as function of forecasting horizon. Each curve corresponds to the different simulation setting: solid* ($T = 250, J = 100$)*, dashed* ($T = 1000, J = 250$)*, dotted* ($T = 2000, J = 1000$)

diction's evaluation. The forecasting of the surface is utilized by forecasting the factors $\widehat{Z}_t$, see Section 2.5.4, and multiplying them with loading functions $\widehat{m}$. The out-of-sample error is considered according to (3.4).

For the comparison between in-sample fit and out-of-sample fit we refer to Figure 3.10. The plot shows $5\%, 25\%, 50\%, 75\%, 95\%$ quantiles of the error. One may clearly observe that out-of-sample fit highly exceeds the in-sample one for each presented quantile. The similar observation applies to the variation of the error. The error tends to increase initially with the forecasting horizon and stabilize afterwards. This feature comes from autoregressive property of the VAR process.

The first impression on Figure 3.10 maybe be discouraging for the application of the DSFM. The fair comparison, however, needs to be done with the prediction based on the true model, from which the data are generated. Here we perform such a study by looking at the forecasts based on the initially simulated Z_t. We consider two cases. First, only time series Z_t is taken into account, which means that the VAR parameter $\widehat{\theta}$ is estimated from Z_t and on this basis the prediction is reached. In the second case, we assume that θ is known and apply the true model of Z_t in order to get the prediction. The first case we refer as *true* Z_t and the second one as *true* θ.

The mean forecasting errors are plotted in Figure 3.11 for three different

simulation settings. One may clearly observe that all cases follow the same path. The difference between the true Z_t and the DSFM is practically indistinguishable. The error in case of true θ' is slightly smaller but it vanishes while the number of observation is increasing. This observations are inline with Theorem 2.9.4 and its implication discussed in Section 2.9.2.

Although in Figure 3.11 the true Z_t and the DSFM errors look practically same there is a discrepancy between them. In order to show it we plot in Figure 3.12 the relative absolute difference between them, where the error of the true Z_t is treated as a benchmark. To make the statement more clear we consider $|MISE_{\widehat{Z}_t} - MISE_{Z_t}|/MISE_{Z_t}$. In the three settings the difference is decreasing with the forecasting horizon. However, as the number of observation is growing the error based on $\widehat{Z}_t$ is converging the to error based on the true Z_t. This fact conveys the justification of the inference on estimated $\widehat{Z}_t$.

3.5 Design Experiments

The simulations in the previous sections study the performance of the model, when parameters describing the data, like T or J, and estimation procedure, like K, vary. In this section we keep the constant setting namely $T = 1000$, $J = 250$, $H = 150$, $\sigma = 0.025$. The factor time series and loading functions are also fixed like described in Section 3.1. The main aim here is to give a short overview on the model's performance, when the design changes.

We consider two types of data design. In Section 3.5.1 some results are presented in case when the data is uniformly spread throughout the space. It does not necessarily mean that X_{tj} has continuous uniform distribution, since we impose the additional restrictions. What we mean is that the probability of obtaining a data point in each possible location is equal. In Section 3.5.2 we neglect this condition and study the case with the skewed distribution of design points.

In order to show some features of the DSFM we present here also a factorial method based on the fPCA, see Section 2.8.2. We enforce an initial spatial bias by estimating the surface on the regular grid of 20 equidistance points in each direction using kth nearest neighbor estimator with 5 neighbors points. For the details on kth nearest neighbors estimator we refer to Härdle (1990). Our aim is not performing a comparative analysis. We rather want to confront some behavior of the DSFM with the behavior of other factorial method, in this case based on functional data analysis. Therefore, we find irrelevant any direct optimization of the parameters of this method.

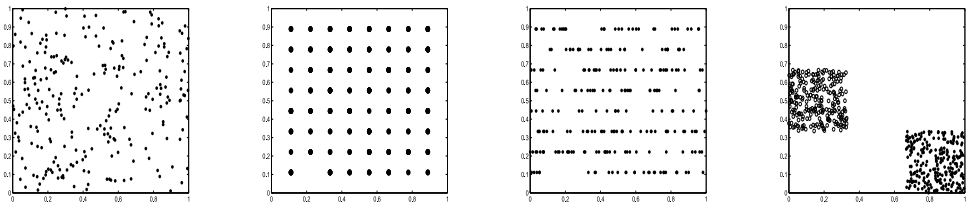

Figure 3.13: *Graphical representation of data designs used in simulation studies. The simulated points come with equal probabilities.*

3.5.1 Uniform Design

In Figure 3.13 we present four different data designs which we consider in this section. The most left panel the data comes from the unform distribution on the unit square. This design corresponds to the experiments studied in Section 3.2 - Section 3.4. In the second type of the experiments the data is observed on some regular 8×8 grid. First the grid point is chosen with equal probability and then the values for this particular grid point. The number of observations in each point has to differ and there could be some points without any observation, see the corresponding panel in Figure 3.13. The third case considers 8 discrete points in one direction and uniform distribution in the second direction. This results in the string structure of the data design as could be seen in the third left panel in Figure 3.13. In the fourth situation (very right panel) the data in each period is uniformly distributed on one of the nine smaller squares. First with equal probability the square is selected and the two dimensional uniform distribution is scaled to this square.

In-sample MSE and MISE of the DSFM and the fPCA are presented in Table 3.5. MSE of the DSFM is here always smaller. One may expect this fact since the DSFM minimize this kind of error in the estimation directly. However it is not the case when one considers $MISE$. Some initial smoothing, although biased, may yield to better fit in the functional sense, when the decomposition of the covariance operator is applied. Focusing on the local structure like in the DSFM may induce inferior global features.

Out-of-sample $MISE$ is presented in Figure 3.14. The difference between the DSFM and the fPCA is hardly distinguishable in all cases. The behavior of the error resembles the behavior described in Section 3.4 with one exception, namely with the squares design. Here in-sample $MISE$ of the fPCA is smaller than $MISE$ of the DSFM, which one can clearly see from the bottom right panel of Figure 3.14 and Table 3.5. The most prominent contrast can be found in smaller out-of-sample fit than in-sample fit. This unusual property could be explained by the fact that only 1/9 of the whole space is

covered in each time period. The long term prediction converge to the mean surface while in case of the in-sample fit the model tries to reproduce the data in one particular square possibly well and neglects the other parts of the surface. This may yield to inferior $MISE$.

design	DSFM MSE	fPCA MSE	DSFM $MISE$	fPCA $MISE$
uniform	$6.170 \cdot 10^{-4}$	$4.921 \cdot 10^{-3}$	$8.299 \cdot 10^{-6}$	$2.431 \cdot 10^{-5}$
points	$6.170 \cdot 10^{-4}$	$4.387 \cdot 10^{-3}$	$1.198 \cdot 10^{-4}$	$1.029 \cdot 10^{-4}$
strings	$6.171 \cdot 10^{-4}$	$4.730 \cdot 10^{-3}$	$1.598 \cdot 10^{-5}$	$5.739 \cdot 10^{-5}$
squares	$6.623 \cdot 10^{-4}$	$1.702 \cdot 10^{-3}$	$3.919 \cdot 10^{-3}$	$2.400 \cdot 10^{-3}$

Table 3.5: *In-sample errors for different uniform designs in a simulation study.*

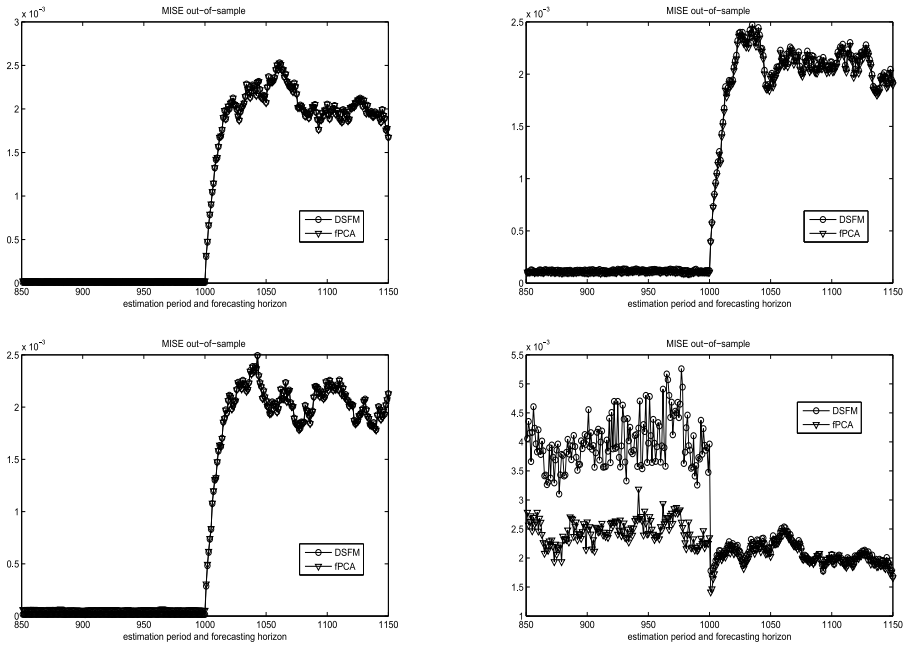

Figure 3.14: *$MISE$ for different data designs. The plot presents comparison between in-sample and out-of-sample fits. Top left panel: uniform. Top right panel: points. Bottom left panel: strings. Bottom right panel: squares. A comparison between the DSFM (circles) and the fPCA (triangles) is also given.*

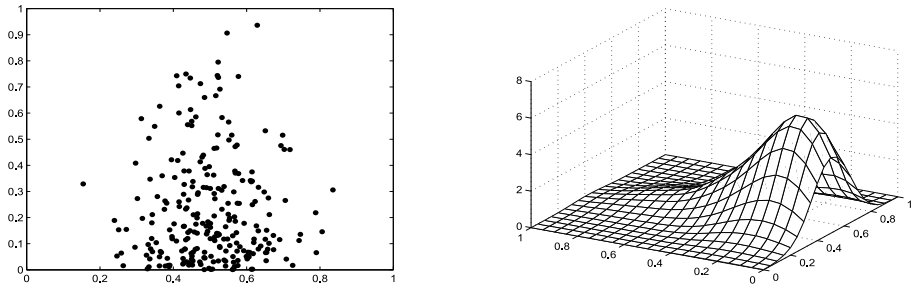

Figure 3.15: *Graphical representation of skewed design. Left panel: example of the design points in one particular time point. Right panel: kernel density estimator based on the pooled data of a single simulation.*

3.5.2 Skewed Design

In the previous section we present some results where the points X_{tj} are equally spread throughout the space. Here we show some issues when the design points are concentrated in a one particular region and come form skewed distribution. The distribution is displayed in Figure 3.15. The left panel shows the typical projection of the data for one particular time point. Note that the certain regions of the space are relatively sparse. In the right panel one may observe the typical empirical data density estimated with kernel density estimator. From this picture one can clearly recognize the way how the design points are generated. In the first direction one takes the symmetric distribution with the mode in 0.5. In the second direction one applies the exponential decay of the probability of obtaining the observation.

The described data design results in some certain difficulties in estimating the DSFM. In the cases discussed up to this moment we use the equidistance knots for estimating the model. Here, however, the model cannot be estimated at all, because the singular matrices appear. For the similar problem in kernel estimation method we refer to the discussion in Section 2.9.1. To overcome this difficulty one may place the knots in the empirical quantiles of the design distribution. For this purpose we transform the data with empirical marginal distribution functions and apply the estimation algorithm to the transformed data. Our transformation avoids setting knots in empty spaces observed in Figure 3.15. One does not suffer the problem with this particular fPCA approach since the kth nearest neighbor smoothing method is chosen.

MSE of the DSFM is $6.277 \cdot 10^{-4}$, which means it is in similar order as in Section 3.5.1. MSE for the fPCA is $2.133 \cdot 10^{-3}$. For $MISE$ we refer

to Figure 3.16. In this case the fPCA has smaller in-sample error than the DSFM. However, the difference of the out-of-sample errors is very small, which is inline with the results of the Section 3.5.1.

The presented study on the design sensitivity illustrates different perspectives of the two discussed methods. As the DSFM minimizes the squared loss, it always gives smaller MSE. In case when for each time point t there is significantly large area without observations, even naive smoothing plus the PCA approach gives smaller in-sample $MISE$. This can be explained by localizing feature of the DSFM, which is in favour of MSE but is penalized when one considers $MISE$. However, for the out-of-sample fit both methods give very similar results.

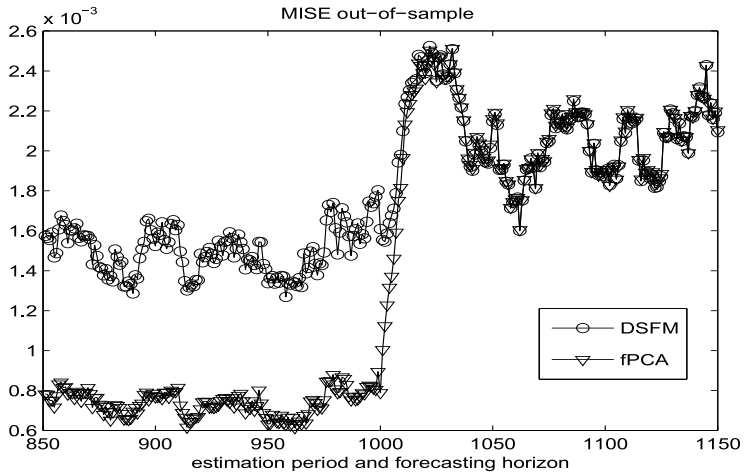

Figure 3.16: *$MISE$ for the skewed data design. The plot presents comparison between in-sample and out-of-sample fits, and between the DSFM (circles) and the fPCA (triangles).*

Chapter 4

Applications

"In theory, there is no difference between theory and practice; In practice, there is."

Chuck Reid

In Chapter 3 the studies on the simulated data are performed. In this chapter we focus on the analysis of the real data. Contrary to the simulation studies the true data generating process is not known and presumably far more complex than the assumed model. Therefore modelling everything is impossible and by assuming a simple model one hopes to obtain a close approximation for further inference. Nevertheless, it would be naive to believe that one particular model always works. Each real life problem requires careful treatment. Even for the same data set two inferential challenges may be solved by two different modelling approaches.

The introduced applications utilized the general idea of factor model (1.2) and the particular attention is paid to the DSFM. We are far from the statement that the DSFM resolves optimally all presented problems. In some cases it may be inferior to traditional approaches, however, this analysis does not aim to provide a golden rule for any type of dynamic issue but rather illustrate how the model can be applied to the real data. It may serve as a motivation for further development of other methods, e.g. parametric methods, or as affirmative tool.

This chapter is structured as follows. In Section 4.1 we present the problems and the data discussed later on. In Section 4.2 we fit the DSFM to different data sets and show the exploratory type of analysis. The inferential point of view is regarded in the next two sections. Section 4.3 discusses the

forecasting power of the model illustrated on the electricity term structure and Section 4.4 focusses on empirical hedging experiment of the knock-out barrier options.

4.1 Practical Issues

This section is focussed on the practical issues of the presented applications. For each problematic we give a motivation, discuss classical approaches and show the data used in the further analysis.

4.1.1 Implied Volatilities

Motivation

Developed financial markets besides primary instruments like stocks and bonds trade extensively secondary instruments like derivatives. A standard example is a plain vanilla European call option, which offers a right to buy an underlying for a predefined strike price K at some future time point. In very similar way the put option gives the right to sell the underlying. Apart from standard plain vanilla contracts it is common to trade exotic options like knock-out barrier options, which offer a payoff only when the price of the underlying does not reach the specified level. One of the challenges of the mathematical modelling is to establish a price and price relations between these financial products.

The origin of financial mathematics for pricing contingent claims goes to Bachelier (1900) and Samuelson (1965). The most prominent and probably most cited paper in quantitative finance is the work of Black and Scholes (1973). Their paper together with further extension of Merton (1973) remains an important benchmark of financial modelling. By applying a replication strategy on the self-financing portfolio one obtains the same payoff as the plain vanilla European option. Consequently by the no-arbitrage arguments this has to be the price of the derivative, which is free from the risk preferences.

Obviously the Black-Scholes model has several assumptions. One may trade infinitesimally divisible assets, there is no difference between lending and borrowing rate r, which stays constant in time and for each horizon. No taxes or short-sale constraints are imposed and the volatility of the underlying remains constant in time. The presented model assumes continuous trading on the time horizon $[0, T]$ and probability space $(\Omega, \mathcal{F}, P)$. The filtration is defined by Wiener process W_t. The price of the tradable asset S_t is a stochastic process given by stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{4.1}$$

where the μ is the constant drift and σ is the volatility. The parameter μ describes the trend of the price evolution and σ the intensity of random deviations from the trend, which are caused by vibrations in price due to e.g. temporary imbalance in supply and demand. In model (4.1) one obtains the price of the European call options as

$$C_t(S_t, \mathrm{K}, r, \tau, \sigma) = S_t \Phi(d_+) - \mathrm{K} e^{-r\tau} \Phi(d_-) \tag{4.2}$$

where $d_\pm = \frac{\log \frac{S_t}{\mathrm{K}} + (r \pm \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$ and $\Phi(\cdot)$ is a cumulative distribution function of standard normal random variable. τ indicates the time left to expiry i.e. $\tau = T - t$. The put option prices can be calculated from Put-Call parity

$$P_t(S_t, \mathrm{K}, r, \tau, \sigma) = C_t(S_t, \mathrm{K}, r, \tau, \sigma) - S_t + \mathrm{K} e^{-r\tau}. \tag{4.3}$$

For the rigorous mathematical treatment of the Black-Scholes model and the derivation of the (4.2) we refer to Musiela and Rutkowski (1997) or Karatzas (1997).

Due to relative simple assumptions and numerical tractability of (4.2) the model achieved great popularity among practitioners. The five parameters of the option price can be grouped into three categories. First, S_t and r may directly obtained from the market data. Of course there is a plethora of possible choices for r, since constant in time risk free interest rate, which reveals flat term structure, does not exist in practice. Second, K and τ are specified in the option contract. While K is a fixed number, τ changes deterministically with time by decreasing linearly to zero through the life time of the option. Finally σ is not observable or specified volatility parameter and has to be estimated from historical prices. It reflects the variability of the asset price. The bigger is uncertainty of the possible asset price change the higher is the call option price. The call option price is in fact an increasing function of σ so there exist one-to-one mapping between option price and the volatility.

The unique mapping from price to volatility leads to the concept of implied volatility (IV), first introduced by Latané and Rendelman (1976). Since it is possible to observe the option price $\widetilde{C}_t$ one may invert (4.2) and find such $\widehat{\sigma}$ that observed price match the Black-Scholes price. Although there is no analytical solution, IV can be efficiently obtained by bisection or Newton-Raphson algorithm.

Since one may observe parallel options with different strikes and maturities it is common to consider the implied volatility surface (IVS), which is

defined as a mapping to $\mathbb{R}^+$

$$(\mathrm{K}, \tau) \rightarrow \widehat{\sigma}_t(\mathrm{K}, \tau).$$

A convenient way of presenting the IVS is to rewrite it as a function of time to maturity and moneyness κ, which is defined in general form as

$$\kappa = m(t, T, S_t, \mathrm{K}, r),$$

where m is the increasing function in K. From now on we will consider the IVS as a function of moneyness κ and time to maturity τ

$$(\kappa, \tau) \rightarrow \widehat{\sigma}_t(\kappa, \tau).$$

The possible choices for moneyness are discussed in Hafner (2004). Here we follow Fengler (2005b) and set the moneyness to forward (future) moneyness $\kappa = \frac{\mathrm{K}}{e^{r\tau} S_t}$.

The particular interest in the IVS comes from its spatial and temporal features. Contrary to the assumptions of the Black-Scholes model, where the volatility is constant, empirical findings show that the IVS reveals a non-flat profile across moneyness (called *smile* or *skew*) and time to maturity. Moreover, it is subjected to random deformations in time. The described features of the IVs can be observed in Figure 4.1, where we plot IVs observed on two different days.

There are several reasons why the IV and in particular the IVS could be important for financial engineering. The IV is often regarded option based estimator for the future realized volatility. It can also serve as market uncertainty indicator. Here we treat the IVS mainly as scaled option prices, which quoted in IVs have the comparable order. Then, the understanding of the IVS dynamics is equivalent to understanding the joint dynamics of the option prices, which may be crucial for pricing, hedging and risk management of some exotic derivatives.

Classical Approach

Since the volatility plays a central role in option pricing, trading and risk management there is a vast stream of literature addressing directly and indirectly the deficiencies of Black-Scholes model. Here we briefly underline two research paths, which approach the problem of smile existence and IVS dynamics. First one considers different underlying dynamics assumption, which results in different option prices. The stochastic differential equation (4.1) is refined for example to

$$dS_t = \mu S_t dt + f(S_t, t, \sigma_t) dW_t, \tag{4.4}$$

where $f(\cdot)$ is some function of spot, time and volatility. A wide class of so called stochastic volatility models treat additionally σ_t as a separate stochastic process. Just to name some of them J. Hull and White (1987) take $f(\cdot) = \sqrt{(\sigma_t)}$ and geometric brownian motion for volatility process, Stein and Stein (1991) take Ornstein-Uhlenbeck process to model volatility and very popular model of Heston (1993) assume mean reverting structure. For different possible specification we refer to, Hagan et al. (2002), Heston and Nandi (2000), Schöbel and Zhu (1999), among others. Reducing in (4.4) $f(\cdot)$ to a function of spot and time drives to another wide class of models. In particular when f is nonparametric one obtains the local volatility model which can replicate the market prices to the arbitrary given precision. Its origin goes to Dupire (1994), Derman and Kani (1994) and achieved great popularity among practitioners. We return to this approach in Section 4.4, where we use it for pricing barrier options in the empirical hedging experiment.

Apart form substituting σ in (4.1) one may consider other nature of randomness. This lead in general to inclusion jumps and to models based on Lévy processes like Merton (1976) or variance gamma model studied in Madan and Seneta (1990) or Carr et al. (2002). A possible combination of stochastic volatility and jumps in one model is proposed by Bates (1996). More comprehensive overview on jump diffusion processes in option pricing can be found in Cont and Tankov (2004). The plethora of stochastic models yield a delicate issue of model choice and model risk discussed in Figlewski and Green (1999), J. C. Hull and Suo (2002), Hirsa et al. (2003), Schoutens et al. (2004). Different approaches may have dissimilar explanation power of the market phenomena, but even if plain vanilla options prices are replicated to similar magnitude, the models may disagree with price of certain exotics, see J. C. Hull and Suo (2002), Detlefsen and Härdle (2007).

The second research path, that we address, is looking on the IVS from statistical perspective. For an empirical investigation on IVS dynamics we refer to Heynen (1994) and Fengler (2005b). A popular approach for its simplification utilize the PCA. Alexander (2001) uses it on the fixed strike volatility deviation from at-the-money volatility. Skiadopoulos et al. (1999) propose bucketing with similar moneyness and maturity, Zhu and Avellaneda (1997) perform the PCA on the term structure of currency options. Common PCA is studied by Fengler et al. (2001) for different maturities across moneyness given on discrete support and Benko et al. (2008) generalize this approach in functional data analysis framework proposing a bootstrap tests for testing equality of eigenvalues and eigenfunctions. Factor model of Section 2.8 are

proposed by Hafner (2004) - parametric model, Cont and Fonseca (2002) - functional PCA approach, Fengler et al. (2007) - DSFM. The evolution of the IVS can be also studied by several possible parameterization like Rosenberg (2000) or Derman (1999), where the skew changes results from evolution of at-the-money volatility.

Data

The data analysis in this chapter is based on the options on the German stock index DAX (ODAX) traded on Deutsche Börse in Frankfurt. DAX is a capital weighted stock index based on the 30 German companies. The companies are selected according to the liquidity and capitalization. The dividends of the companies are reinvested into the index.

The traded options are European vanilla puts and calls. The exchange set the expiry of the option on the third Friday of the particular month and if it is not a trading day then the closest trading day before. During a particular trading day several maturities are available: three nearest successive calendar months, quarterly running contracts expiring in March, June, September or December, semi-annual running contract expiring in June or December. Typically one may trade several different strikes but the smallest difference between the strikes is 50 points. Usually the strikes lays symmetric around the current underlying security.

The option data from Deutsche Börse have two basic forms. The intra-day data records the price of each traded contract together with its characteristics. The daily data reflect the last traded price and if it is more than 15 minutes old or does not reflect the market conditions the official settlement price is established. Based on this data one calculates the Black-Scholes implied volatility.

Although the IV from puts and calls should be theoretically same it is not always the case in the real data. This mismatch could be for example explained by taxation legislative, see Hafner and Wallmeier (2001) or Benko (2006). The traditional procedure to overcome this hurdle is to consider the corrected spot $\widetilde{S}_t$ in such a way that formula (4.3) holds. For the details on this type of corrections we refer to Hafner and Wallmeier (2001) or Fengler (2005b).

The trading convention results in particular dynamic string structure of the IV data. This can be clearly observed in Figure 4.1. The bottom solid lines indicates the observed maturities, which appear only in few discrete points. On two different days the 'strings' are observed in different positions. This feature can be recovered from the comparison between the left and right panel of Figure 4.1 and comes from the fact that the volatility string

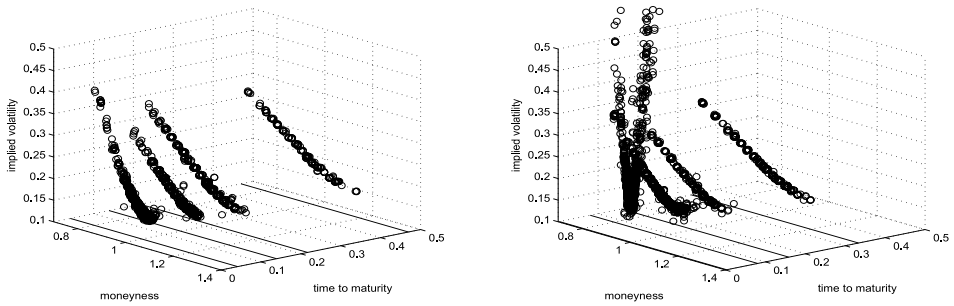

Figure 4.1: *Typical IV data design on two different days. In the maturity direction observations appear in the discrete points for each particular day. Bottom solid lines indicate the observed maturities, which move towards the expiry. Left panel: observations on July 1st, 2004, $J_t = 5606$. Right panel: observations on August 19th, 2004, $J_t = 8152$.*

shift towards expiry. The figure shows also that in the moneyness direction the observation lie relatively dense since many possible strikes are available and the underlying fluctuates. Moreover, for the option market it could be clearly observed that the frequency of the trades are not uniform. One can find significantly more market activities for the options closer to expiry and at-the-money ($\kappa \approx 1$). This fact is illustrated in Figure 4.2, where the kernel density estimate of the design points is plotted. The detailed discussion on the issues of irregular design of the IV data can be found in Fengler (2005b).

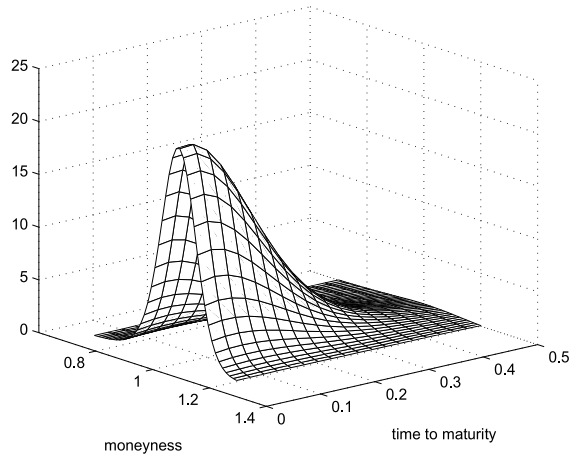

Figure 4.2: *Kernel density estimator of the design points from July 1st, 2004 to June 29th, 2005.*

In Section 4.2.1 we analyze the data observed from July 1st, 2004 to June 29th, 2005. The one year period corresponds to the financial regulatory requirements. The data are taken from Financial and Economic Data Center of Humboldt-Universität zu Berlin. The IV data are regressed on the two-dimensional space of future moneyness and time to maturity, denoted by $(\kappa_t, \tau_t)^\top$. For our application we chose r_t as a daily EURIBOR rates, taken from Ecowin Reuters database, linearly interpolated between observed maturities in order to match the maturities of the options. The time-to-maturity of the options is measured in years. We take all trades with $10/365 < \tau < 0.5$. We limit also the moneyness range to $\kappa \in [0.7, 1.2]$.

Section 4.4 concentrates on empirical hedging analysis with respect to the most common IVS movement. For this aim we take DAX settlement option prices traded at EUREX from January 3rd, 2000 till June 30th, 2004, which gives altogether 1135 trading days. We process the data eliminating implied volatilities bigger than 80% and maturities smaller than 10 days. The possible arbitrage violations in the settlement data are removed by arbitrage free smoothing procedure described in Fengler (2005a). This proceeding has a crucial meaning since otherwise one may obtain option prices, that are not economically meaningful. After smoothing the IVS data are converted to the regularized grid of moneyness and maturity. For the computation of the option prices zero rates from Bloomberg are used. We linearly interpolate them to get rates for any arbitrary maturity, see Dumas et al. (1998).

4.1.2 CO_2 Emission Allowance

Motivation

In January 2005 the EU-wide CO_2 emissions trading system (EU-ETS) has formally entered into operation. Within the new trading system, the right to emit a particular amount of CO_2 becomes a tradable commodity - called EU Allowances (EUAs) - and affected companies, traders and investors will face new strategic challenges. The EU-ETS requires a cap-and-trade program whereby the right to emit a particular amount of CO_2 becomes a tradable commodity ISI (2003). Since environmental policy has historically been a command-and-control type regulation where companies had to strictly comply with emission standards, the new trading system represents a shift in paradigms. After an initial pilot trading period from 2005-2007, in 2008 there is a new allocation plan in each of the countries and the first Kyoto-commitment trading period starts lasting until 2012.

Since failure to submit a sufficient amount of allowances results in sanction payments per missing ton of CO_2 allowances, the new market forces

companies to hold an adequate number of allowances according to their carbon dioxide output. For example, failure to submit a sufficient amount of allowances results in sanction payments of 40 Euro per missing ton of CO_2 allowances during the pilot period and 100 Euro in the commitment periods. While allowance trading has primarily been applied in the US, the EU-ETS will result in the world's largest greenhouse gas (GHG) emissions trading system. In fact, all combustion installations exceeding 20 MW will be affected by the trading scheme including different kinds of industries like metal, cement, paper, glass etc. as well as refineries or coke ovens.

Classical Approach

The literature on the EU-ETS on price behavior, risk management or hedging with CO_2 spot or future contracts is very sparse. The majority of publications on greenhouse gas emissions assesses the US market where emission trading was already established in the early 1990s. By using industrial organization models they account for changes in parameters of technology Rezek (1999) and electricity demand Schennach (2000) and their impact on the optimal equilibrium price path. There is also a number of empirical investigations on ex-post market price analysis, among them Burtraw (1996) and Ellerman and Montero (1998).

For the European market, Maeda (2001) provides a rather theoretical analysis on banking impacts and forward pricing on the market while Uhrig-Homburg and Wagner (2006) investigate the success chances and optimal design of derivatives on emission allowances. For CO_2 market price simulation studies with respect to changes in market design parameters see e.g. Burtraw et al. (2002), Böhringer and Lange (2005) and Schleich et al. (2006). Finally, Benz and Trück (2006) as well as Paolella and Taschini (2006) provide an econometric analysis on price behavior of allowance prices and investigate different models for the dynamics of short-term price behavior. However, non of the papers takes into account the CO_2 allowance futures market.

Data

For our analysis we use all futures quotes available on the European Energy Exchange (EEX) in Leipzig during the time period from October 4, 2005 to September 29, 2006. Hence, the time period comprises approximately the first year of futures trading at EEX. Spot contracts for EU emission allowances have a contract volume of 1 ton CO_2 and are traded in Euro up to two decimal points. The object of a European Carbon Future contract is the delivery of EU emission allowances for the first period of three years

beginning on January 1, 2005 or for the second period of five years beginning on January 1, 2008. Hereby, the contract volume amounts to $1,000$ t CO_2 while maturity occurs on the last day of trading of a futures contract, namely the penultimate exchange trading day in the month of November 2006 and 2007 for the pilot period and November 2008 to 2012 for the Kyoto commitment period. For every futures contract a settlement price in accordance with the current market price is established on a daily basis. According to a daily profit and loss balancing (variation margin), the change in the value of a futures position is credited to the trading participant in cash or debited with him in cash. Delivery of the EU emission allowances will be carried out two settlement days after maturity of a futures contract, i.e. on the first settlement day in December of the corresponding year.

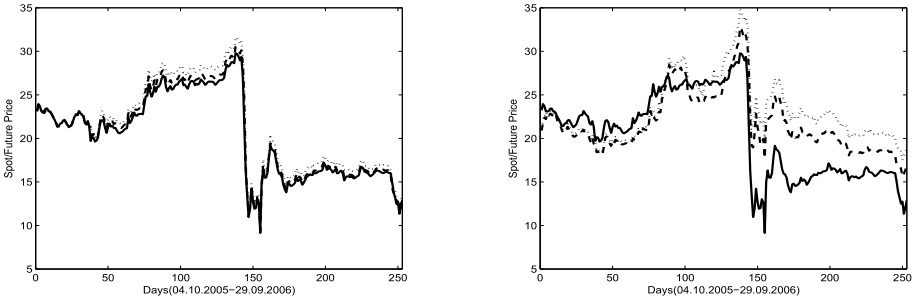

Figure 4.3: *EEX emission allowance spot and futures rices for Oct 4, 2005 - September 29, 2006. Left panel: spot (solid) and future prices with delivery in 2006 (dashed) and 2007 (dotted). Right panel: spot (solid) and future prices with delivery in 2009 (dashed) and 2012 (dotted)*

While spot trading has started already in January 2005, when the EU-wide CO_2 emissions trading system entered into operation, future contracts have been traded only since October 2005. To investigate the future allowance prices, we consider the time period starting from October 4, 2005 until September 29, 2006, when both spot and futures were traded at EEX. Figure 4.3 in the left panel displays spot and emission allowance future prices for delivery in 2006 and 2007 while the right panel shows the futures prices for delivery in 2009 and 2012 for the considered time period. At the commencement of the trading period, prices initially fell due to a quite mild climate and high supply of wind energy from Scandinavia and North Germany. However, at the end of January an extreme cold snap and constant high UK gas and oil prices, compared to relatively low coal prices, led to a drastically price increase within the next months. This effect was boosted by an extremely dry summer in the southwest of Europe. The consequence of

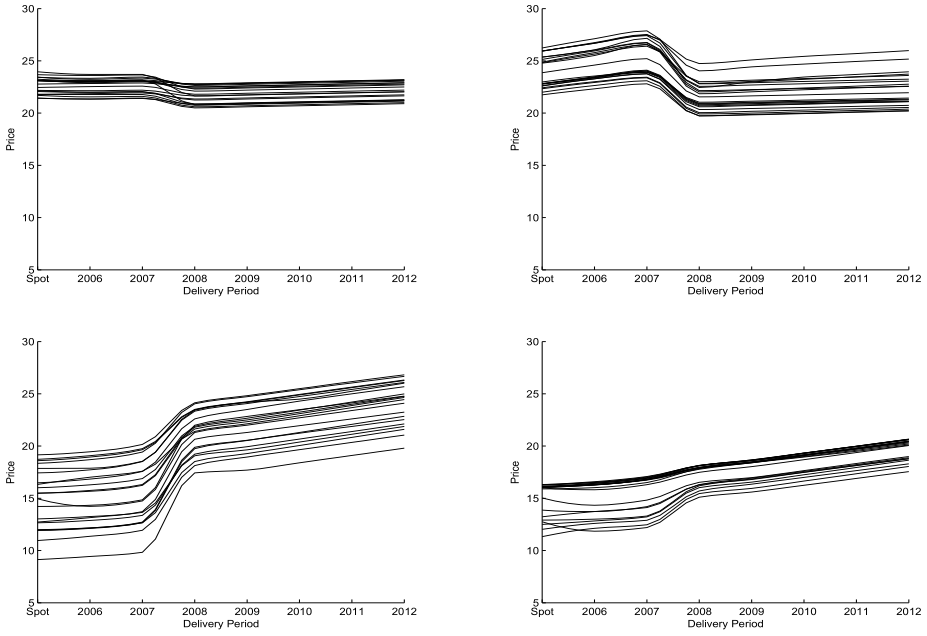

Figure 4.4: *Term structure for spot and futures prices for each day of October 4 - 31, 2005 (upper left panel), January 1 - 31, 2006 (upper right panel), May 1 - 31, 2006 (lower left panel) and September 1 - 29, 2006 (lower right panel).*

the high temperature and absence rainfall was to prevent full utilization of hydraulic plants, especially in Spain. Additionally, the lack of cooling water for nuclear power plants led to a higher power plant utilization and therefore increased the demand for CO_2 permits. Prices peaked on July 11 with 29.21 Euro but fell back to a level of approximately 22 Euro in August, remaining there until the end of 2005. Again, the beginning of an extremely cold winter in January 2006 led to a substantial increase in allowance prices. While temperature remained cold also in April 2006, the so far highest price could be observed on April 18 with 29.78 Euro. Shortly after this news spread that many countries participating in the EU-ETS had given their industries so generous emission caps that there were no need for them to reduce emissions. Prices fell dramatically within three weeks from 29.37 Euro on April 24 to 9.13 Euro on May 12. Until the end of May a renewed increase of spot prices to approximately 18 Euro could be observed until the end of May. In June prices fell to approximately 14 Euro in June and remained between 14 and 17 Euro until mid September. Finally, in September during the last two weeks of the considered period prices fell approximately 12 Euro.

Figure 4.4 displays the term structure of emission allowance spot futures prices with yearly maturities from November 2006 to November 2012. For each trading day in October 2005, January 2006, May 2006 and September 2006 the observed spot and futures prices are connected by a smoothed line, yielding between 20 (October 2005) and 23 lines (May 2006) in each of the subfigures. We find that the term structure of futures prices is dynamic and shows quite different behavior through time. During the initial trading period in October 2005 futures prices both for the pilot and Kyoto period were slightly below current spot prices. While there was a quite flat term structure for the pilot period, a slightly increasing term structure of futures prices could be observed for the Kyoto commitment period. In January 2006, for the pilot period an increasing term structure can be observed while the term structure for the Kyoto period is only slightly increasing. Futures prices for the commitment period are still below the spot price and futures prices of the pilot period. In May 2006, after the news of overallocation of emission rights in a number of European countries was spread, futures prices for the Kyoto period are clearly higher than the spot and period 1 futures prices. A similar relationship between spot and futures prices can be found for the last month of the examined period. In September 2006, an increasing term structure can be observed and futures prices for the Kyoto period are still above the spot price and period period future prices.

4.1.3 Functional Magnetic Resonance Imaging

Motivation

The human brain is the most mysterious and complex part of the body. Understanding its structure and mechanism arise great fascination among the scientific community. There are several techniques which enable to recover some part of the neural activity. The very popular one is functional magnetic resonance imaging (fMRI).

The fMRI is a non-invasive method for recovering the brain mechanics. Changes in neural activity infer changes of the relative concentration of oxygenated and deoxygenated blood. Since deoxyhemoglobin is paramagnetic unlike oxygenated hemoglobin by measuring the frequency of electromagnetic waves one obtains the map of so called blood-oxygen-level-dependent (BOLD) signals. In the typical neural-imaging experiment the subjects are put for approx. 10 minutes in the strong magnetic field and given a stimuli. Since the higher oxygenation level is associated with higher neural activity necessary for responding to the impulse one may localize brain's regions responsible for particular activity.

As a result of the experiment one obtains series of 3 dimensional images on relatively high resolution. The scans are taken every 1-4 seconds and the image cubes contain both brain and out-of-brain observation. This spatial and temporal structure is of high interest for statisticians who apply their methods in order to recover interesting patterns.

As any tool the fMRI has its limitation. It is only indirect measure since BOLD signal is not perfectly correlated with neural events. Additionally, the contrast of the signals in two locations cannot be to the measure of the contrast of neural response. More details on interpreting the BOLD signal can be found in Logothetis and Wandell (2004).

Classical Approach

Ogawa's work, just to list few papers Ogawa, Lee, Kay, and Tank (1990), Ogawa, Lee, Nayak, and Glynn (1990) or Ogawa and Lee (1990), caused great interest in human BOLD fMRI. Further works of Kwong et al. (1992), Ogawa et al. (1992), Bandettini et al. (1992) initiated great amount of scientific fMRI publications.

A lot of statistical research in many different directions have been done in order to explain the fMRI signal. One stream of research goes to dynamic component analysis as described in Thirion and Faugeras (2003). In this approach the BOLD signal is explain as a linear combination of temporal and spatial components, similar to the setting of Section 2.8. For this aim one may apply the PCA to extract eigenimages like Friston et al. (1993) or Petersson et al. (1999). Friston et al. (2000) propose nonlinear PCA and Thirion and Faugeras (2003) consider kernel PCA. Very similar spirit is followed by independent component analysis (ICA), where one tries to extract independent signals, see Hyvärinen et al. (2001). ICA for fMRI is studied by Bell and Sejnowski (1995), McKeown et al. (1998) for components independent in spatial domain or Calhoun et al. (2001) for components independent in time domain. Lukic et al. (2002) consider the robust ICA and Liao et al. (2006) discuss motion correction ICA for minimizing the head motion-induced signal variations

Another methods fit a prior model to the data. For example Glover (1999) and Worsley et al. (2002) assume the shape of hemodynamic response. However, the shape of response can vary across the subjects and regions as indicated in Miezin et al. (2000), Svensén et al. (2000) or Duann et al. (2002). The very popular method is statistical parametric mapping, see Friston et al. (2007), which assume univariate regression model for each voxel. Based on this initial linear fit Tabelow et al. (2006) suggest using propagation-separation procedure for locating spatial information. Non-linear more so-

phisticated model is proposed by Purdon et al. (2001).

Apart from described techniques other statistical methods are proposed to study the structure of fMRI data, canonical variate analysis Friston, Poline, et al. (1996), multidimensional scaling Friston, Frith, et al. (1996), partial least squares McIntosh et al. (1996). Clustering based method are considered in Baune and Sommer (1999), Meyer and Chinrungrueng (2005) or Filzmoser et al. (1999). Qui and Lane (2006) discuss the usage of support vector machines to study the communication behaviors. In Section 4.2.3 we illustrate how the DSFM could be applied to fMRI data.

Data

For our illustration in Section 4.2.3 we study the real fMRI data. A subject was scanned at the Max-Planck Institut für Kognitions-und Neurowissenschaften Leipzig on a 3.0 Tesla Siemens TRIO scanner using a standard head coil. Standard gradient-echo echoplanar imaging was used to measure BOLD contrast for functional imaging. The scans were taken every 2 seconds on the resolution of $3 \times 3 \times 2mm^3$ with 1 mm gap between the slices. During the passive experiment the subject was shown 3 types of objects (bench, phone and motorbike), rotated around at randomly changing axis, for 4 seconds. Objects rotated around a randomly changing axis at a constant angular velocity of 1.5 cycle/sec. The scanning phase was followed by relaxation phase of 6-10 seconds. Each stimuli were shown 16 times in pseudo-randomized order.

As result one obtains the series of 290 $64 \times 64 \times 30$ images. An example of a measurement in one particular time point is presented in Figure 4.5 for 15 different slices. Since the image is an 3 dimensional cube one may cut it in 3 different directions. Other cuts are presented in Figure 4.6. From both pictures it is clear to observe that the strongest signal comes from the body's part of the image since some physiological shapes are recognizable.

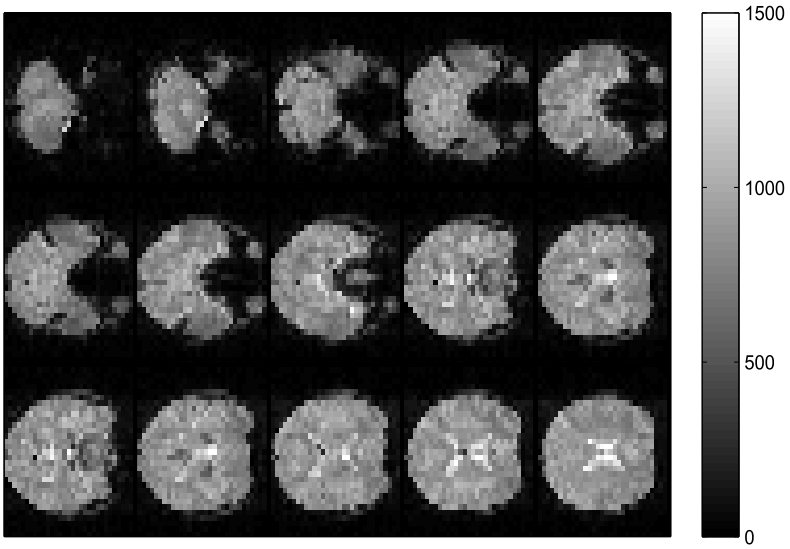

Figure 4.5: *Typical fMRI image in one particular time point. The figure presents* 15 *parallel horizontal images. The brightness corresponds to the strength of the observed BOLD signals.*

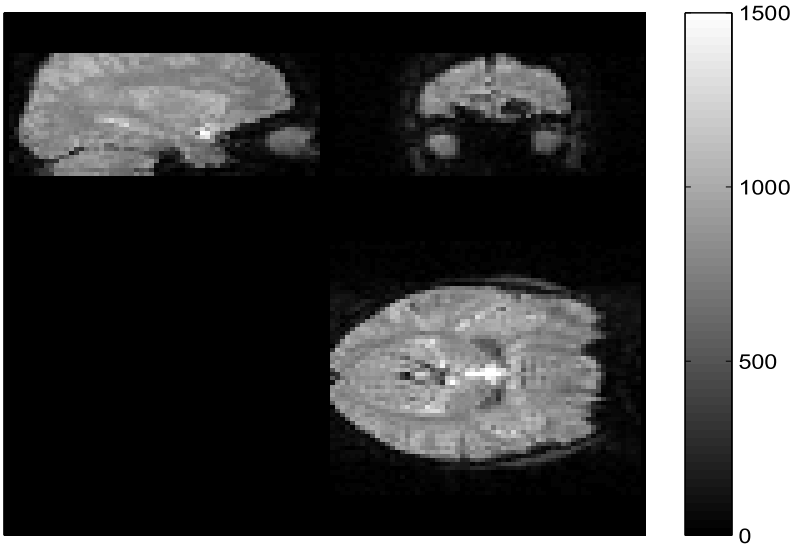

Figure 4.6: *Typical fMRI image in one particular time point. The figure presents side, front and top view of one particular slice. The brightness corresponds to the strength of the observed BOLD signals.*

4.1.4 Electricity Futures

Motivation

In the last two decades dramatic change of the structure of the electricity business have taken place. Traditionally the sufficient supply of the electric power were achieved by centralized regulations. Recently the monopolistic structure have been replaced by competitive markets, where consumers are able to chose their providers. The liberalization of electricity market has led to the development of new financial instruments for electricity. In particular the power exchanges have been organized, which offer spot and derivative contracts. Although main deals are made through long and medium term bilateral contracts the existence of exchange markets promotes competition and serves as source of market information.

The emergence of competitive markets resulted in new risk management issues for the electricity consumers and producers. It is vital for a volumetric risk but the extreme price volatility requires particularly careful financial risk management. It is worth to mention that since the electricity is not a storable good some additional characteristics, like delivery period, has to be present in the contract and therefore they are also taken into account in the derivative contracts specification.

Classical Approach

Modelling for electricity markets may be considered from several perspectives. Weron (2006) propose to classify the models into production cost models, equilibrium approaches, fundamental models, quantitative approaches, statistical approaches and artificial intelligence-based techniques.

Production cost models try to simulate the operation of generating units aiming to satisfy demand at the minimum costs. They are studied in Angelus (2001) or Wood and Wollenberg (2006) . Equilibrium or game theory approach may be regarded as a generalization of production cost model and aim to catch strategic behavior. Cournot-Nash framework for electricity is discussed in Borenstein et al. (1999) or Ventosa et al. (2005) and modelling the price as the equilibrium of companies bidding with supply curve into the whole sale market is given in Green and Newbery (1992), Hinz (2003), Hobbs et al. (2000) among others. The fundamental models incorporate the basic physical and economical factors (loads, weather conditions etc.) present in production and trading of electricity, see Bunn (2004), Kirschen (2003), Stoft (2002). For the derivative pricing goal one adopts from financial literature the quantitative models, which try to mimic the statistical properties of electricity prices. The electricity risk premia is studied in Botterud et al. (2002)

or Longstaff and Wang (2004) and the spot-forward relationship in Geman and Roncoroni (2006). Methods based on the the statistics generated by market activity belong to the class of statistical models. In particular there are time series methods for load and price forecasting like Nowicka-Zagrajek and Weron (2002) or Nogales et al. (2002). In the class of artificial intelligence based models neural networks have received the most attention, see Gao et al. (2000), Zhang and Dong (2001), Szkuta et al. (1999) among others.

More detailed description of the models can be found in Weron (2006). This work describes as well power markets in many countries, their stylized facts and gives several case studies.

Data

The Nordic commodity market for electricity, known as Nord Pool was a first international power exchange. Its origin goes to the deregulation of the electricity sector in Norway and after Sweden (1996), Finland (1998) and Denmark (2000) jointed the common electricity market it is fair to talk about power exchange for the Nordic region. The exchange offers spot trading for physical delivery of the electricity and financial market with futures, forwards (up to three years ahead), options and contracts for differences (contract for price area differentials). The future and forward contracts are written on the arithmetic average of the system price of the given time interval. The contracts refer to delivery of 1 MW during each hour for a delivery period. None of the contract is traded during the delivery period and the market settlement is done on the daily basis. The forward contracts have settlement only in delivery period while the futures contracts have in delivery and during the trading period.

The futures contracts have the shortest delivery period. One may trade the 24 hours contracts within the nearest week, weekly futures with 168 hours delivery 4-8 weeks before, four weeks delivery period contracts. The forward contracts are divided to seasonal contracts - late winter (January 1st - April 30th), summer (May 1st - September 30th), early winter (October 30 - December 31st). All 3 seasonal contracts are available for the next two years. Apart form that yearly forwards are traded.

Our data set contains prices of the electricity future and forwards traded on the Nord Pool from January 4th, 1999 to December 30th, 2004, which is six years of data. For the analysis we consider the term structure with maturities bigger than 7 days and smaller than 2 years. The prices are mainly quoted in NOK/MWh, however for the contracts, which have delivery period in 2006, the prices are given in EUR. In order to unify the currency we recalculate all prices to NOK using spot exchange rates from Reuters EcoWin database.

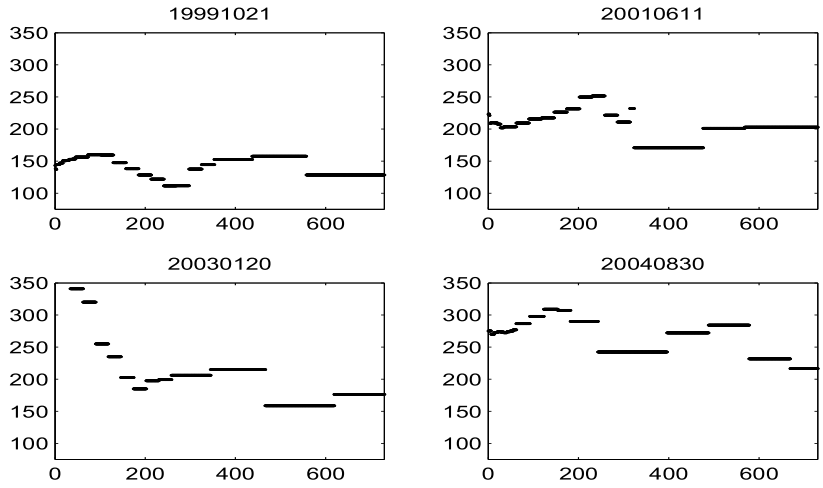

Figure 4.7: *Term structure of the electricity prices observed on four different days: October 21st, 1999, June 11th, 2001, January 20th, 2003 and August 30th, 2004. The price is given in NOK/MWh and maturity is displayed in days.*

The missing data are replaced with the mean of two nearest contracts so that we obtain the prices for delivery on the whole regarded maturity space. In case when the futures and forward contracts overlap for some delivery period we take the future contract prices on this period.

The term structure for the four different days across the sample are displayed in Figure 4.7. Although only one price is quoted it corresponds to some delivery period and therefore the term structure is a piecewise constant curve. The delivery periods shorten near the expiry, which results in more split curve and higher variation for smaller maturities, see Koekebakker and Ollmar (2005). As it could be clearly seen, in four different days the curve exhibit different shapes, which suggest that the term structure of electricity prices is a highly dynamic object.

Moreover, it could be observed that it forms a sort of sinusoidal shape with period approximately 1 year. It is well known that electricity demand exhibits seasonal fluctuations, which mostly arise due to changing climate conditions, like temperature and the number of daylight hours. In some countries also the supply side shows seasonal variations in output. Hydro units, for example, are heavily dependent on precipitation and snow melting, which varies from season to season. These seasonal fluctuations in demand and supply translate into seasonal behavior of electricity spot prices, futures and forwards prices in particular.

4.2 Fitting the Model

This section shows the DSFM fitting to the data presented in the previous section. For this aim one has to specify the size of the model L and estimation characteristics. In case of kernel method it is a type of the kernel function and bandwidths. For the series estimator one has to chose the basis function and, in particular for the B-splines, number of knots and knots location have to be specified.

The estimation results are summarized by time series $\widehat{Z}_t$ and functions $\widehat{m}_l$. The most convenient way for presenting these complex objects is graphical representation, which facilitate the practical interpretation. Although it gives relatively intuitive picture for one and two dimensional functions $\widehat{m}_l$, displaying the functions of three arguments, like in fMRI example, have some natural burdens.

4.2.1 Implied Volatilities

The structure of the IV data, described already in Section 4.1.1, requires a careful treatment. Apart from the dynamic degeneration, one may also observe nonuniform frequency of the trades with significantly greater market activities for the options closer to expiry or at-the-money. To avoid the problems with the highly skewed empirical distribution of $X_t = (\kappa_t, \tau_t)$, we transformed the initial space $[0.7, 1.2] \times [0.03, 0.5]$ to $[0, 1]^2$ by using the marginal empirical distribution functions. We applied the estimation algorithm to the transformed space, and then transformed back the results to the original space. Similar idea has been already presented in Section 3.5.2.

In order to obtain the dynamics of the IVS we apply the series estimator. Based on the computed values of $RV(L)$, which are given in Table 4.1 for various L, we choose $L = 2$. Table 4.1 indicates that the third, fourth and fifth factor make only a small improvement in the fit.

No. Factors	$1 - RV(L)$
$L = 1$	0.848
$L = 2$	0.969
$L = 3$	0.976
$L = 4$	0.978
$L = 5$	0.980

Table 4.1: *Proportion of the explained variation by the models with* $L = 1, \ldots, 5$ *dynamic factors.*

knots		order		$1 - RV(2)$
moneyness	maturity	moneyness	maturity	
15	10	4	3	0.974
10	10	4	3	0.972
10	5	4	3	0.969
5	5	4	3	0.961
15	10	3	3	0.972
10	10	3	3	0.969
10	5	3	3	0.965
5	5	3	3	0.951
15	10	3	2	0.971
10	10	3	2	0.968
10	5	3	2	0.967
5	5	3	2	0.949

Table 4.2: *Proportion of the explained variation by the models when $L = 2$ for different numbers of knots and different orders of splines.*

We use tensor B-splines that are cubic in the moneyness and quadratic in the maturity direction. In the transformed space we place 10×5 knots – 10 in the moneyness and 5 in the maturity direction. We find that the results are not sensitive to the choice of the number of knots and the orders of splines, see Table 4.2. This is inline with the simulation study from Section 3.3.4.

We chose $\widehat{m}_1$ and $\widehat{m}_2$ that are orthogonal to each other in $L_2[0,1]^2$ in such a way that $\sum_{t=1}^{T} \widehat{Z}_{t,1}^2$ is maximized, as is described in Fengler et al. (2007).

The estimated functions $\widehat{m}_1$ and $\widehat{m}_2$ are plotted in Figure 4.8 in the transformed estimation space. The intercept function $\widehat{m}_0$ is almost flat around zero, thus is not given. By construction, $\widehat{m}_0 + \widehat{Z}_{t,1}\widehat{m}_1$ explain the principal movements of the surface. It was observed by Cont and Fonseca (2002) and Fengler et al. (2007) that most dominant innovations of the entire surface are parallel level shifts. Note that VDAX is an estimated at-the-money IV for an option with 45 days to maturity, and thus indicates up-and-down shifts. The left panel of Figure 4.9 shows the values of VDAX together with $\widehat{m}_0(X_{t,0}) + \widehat{Z}_{t,1}\widehat{m}_1(X_{t,0})$, where $X_{t,0}$ is the moneyness and maturity corresponding to an option at-the-money with 45 days to maturity. The right panel of Figure 4.9 depicts the factor $\widehat{Z}_t$, where one can find that $\widehat{Z}_t$ shows almost the same dynamic behavior as the index VDAX. This similarity supports that DSFM catches leading dynamic effects successfully. Obviously the model in its full setting explains other effects, such as skew or term structure changes, which are not explicitly stated here.

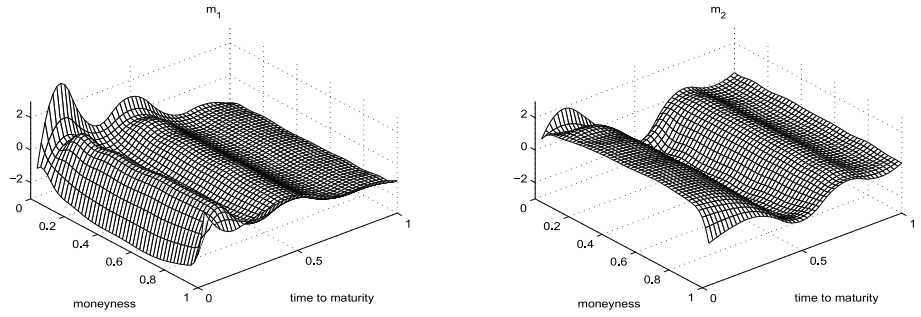

Figure 4.8: *Estimated functions $\widehat{m}_l$ for the ODAX IV data from July 1st, 2005 to June 29th, 2005.*

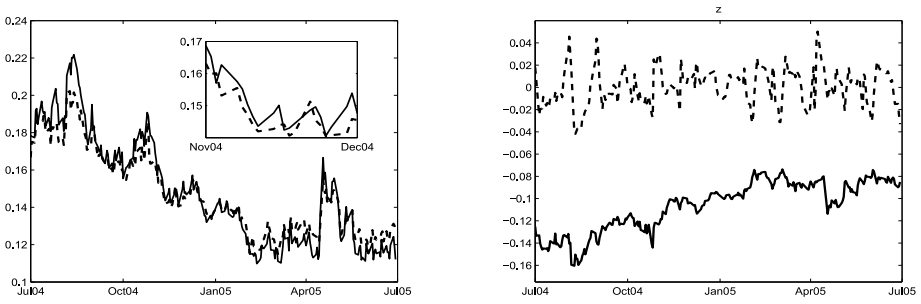

Figure 4.9: *Left panel: VDAX from July 1st, 2004 to June 6th, 2005 (solid) and the dynamics of the corresponding IV given by the sub-model $\widehat{m}_0+\widehat{Z}_{t,1}\widehat{m}_1$ (dashed). Right panel: The obtained time series $\widehat{Z}_t$ on the ODAX IV data from July 1st, 2004 to June 6th, 2005. The solid line represents $\widehat{Z}_{t1}$, the dashed line $\widehat{Z}_{t2}$.*

Statistical analysis on the evolution of a high-dimensional system ruling the option prices can be simplified to a low-dimensional analysis of the $\widehat{Z}_t$. In particular, as our theory in Section 2.9.2 and the simulation results in Sections 3.2 and 3.4 assert, the inference based on the $\widehat{Z}_t$ is well justified in the VAR context. To select a VAR model we computed the Schwarz (SC), the Hannan-Quinn (HQ) and the Akaike criterion, as given in Table 4.3. One can find that SC and HQ suggest a VAR(1) process, while AIC selects VAR(2). The parameter estimates for each selected model are given in Table 4.4. The inverse roots of the characteristic polynomial lie inside the unit circle, so the specified models satisfy the stationarity condition. For each of VAR(1) and VAR(2) models, we conducted a portmanteau test for the hypothesis that the autocorrelations of the error term at lags up to 12 are all zero, and also a series of LM tests, each of which tests whether the autocorrelation at a

order	AIC	SC	HQ
1	-14.06	-13.98*	-14.03*
2	-14.07*	-13.93	-14.02
3	-14.06	-13.86	-13.98
4	-14.06	-13.81	-13.96
5	-14.07	-13.76	-13.95

Table 4.3: *VAR model selection criteria. The smallest value for each criterion is marked by (*).*

	VAR(1)			VAR(2)				
	$\widehat{Z}_{t-1,1}$	$\widehat{Z}_{t-1,2}$	const.	$\widehat{Z}_{t-1,1}$	$\widehat{Z}_{t-1,2}$	$\widehat{Z}_{t-2,1}$	$\widehat{Z}_{t-2,2}$	const.
$\widehat{Z}_{t,1}$	0.984	-0.029	-0.001	0.913	-0.025	0.071	-0.004	-0.001
$\widehat{Z}_{t,2}$	0.055	0.739	0.005	0.124	0.880	-0.065	-0.187	0.006

Table 4.4: *Estimated parameters for VAR(1) and VAR(2) models.*

particular lag up to 5 equals zero. Some details on selection of lags for these tests can be found in Hosking (1980, 1981), Brüggemann, Lütkepohl, and Saikkonen (2006) and see also Section 2.5.3 for the tests' description. We found that in any test the null hypothesis was not rejected at 5% level. A closer inspection on the autocorrelations of the residuals, however, revealed that the autocorrelation of $\widehat{Z}_{t2}$ residuals at lag one is slightly significant in the VAR(1) model, see Figure 4.10. But, this effect disappears in the VAR(2) case, see Figure 4.11. Similar analysis of characteristic polynomials, portmanteau and LM tests supported VAR(2) as a successful model for $\widehat{Z}_t$.

4.2.2 CO_2 Emission Allowance

In this section we study the dynamics of the futures prices of the CO_2 emission allowance in the functional form. One may treat each future contract separately and perform some comparative analysis. Such approach is presented in Trück et al. (2006). Here we consider the whole term structure of the futures to better understand the dynamics of the entire system. For this aim we apply the DSFM.

The DSFM applied to the futures on CO_2 allowances has the following form: let Y_{tj} be a price of the future contract observed in time t and with time to expiry X_{tj}. The index j counts the contracts on a one particular day. For the estimation we apply the kernel algorithm described in Section 2.9.1. For our purpose we choose quartic kernel and set the bandwidths $h = 2$.

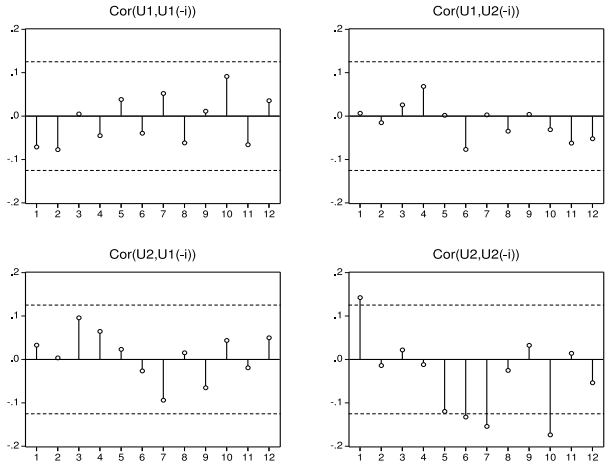

Figure 4.10: *Cross-autocorrelogram for the VAR(1) residuals. The dashed line-bounds indicate* $\pm 2 \times$ *(standard deviations), which correspond to an approximate* 95% *confidence bound.*

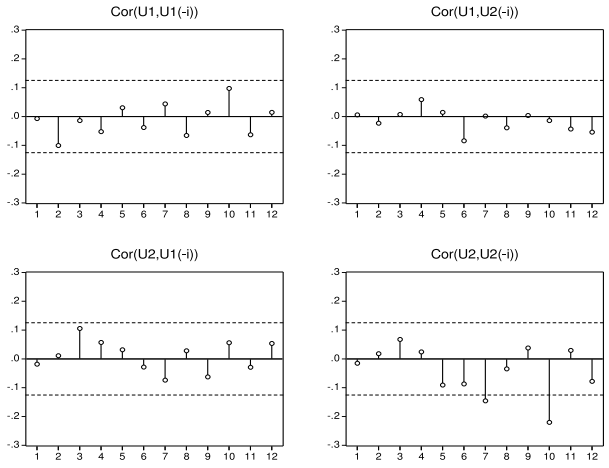

Figure 4.11: *Cross-autocorrelogram for the VAR(2) residuals. The dashed line-bounds indicate* $\pm 2 \times$ *(standard deviations), which correspond to an approximate* 95% *confidence bound.*

No. Factors	$1 - RV(L)$
L=1	0.8755
L=2	0.9860
L=3	0.9881

Table 4.5: *Explained variance for DSFM applied to CO_2 allowance future prices.*

For the choice of L we calculate the $1 - RV(L)$ for different values of L. In Table 4.5 we present the estimation results for $L = 1, 2, 3$. The inclusion of the third function only slightly improve the explanatory power of the fit and therefore from now on we operate only on the model with 2 basis functions.

The estimated functions $\widehat{m}$ and time series $\widehat{Z}_t$ are plotted in Figure 4.12. The first function is relatively flat and it can be interpreted as level changes of the whole term structure. The second function has very pronounced break near time to expiry 2 years. It is exactly the place of separation between the pilot period and Kyoto period. The form of the function reflects different behavior of the future prices before and after Kyoto commitment period. Before the period the function is significantly positive while after it is negative. Therefore, the positive $\widehat{Z}_{t2}$ reflects the higher prices of pilot futures, like in January 2006, and negative $\widehat{Z}_{t2}$ corresponds to higher prices in Kyoto period, like in May 2006. $\widehat{Z}_{t2}$ close to 0 results in relatively flat term structure i.e. no visible break between pilot and Kyoto periods.

To complete our analysis we plot $\widehat{Z}_{t1}$ together with the spot prices (left panel) and $\widehat{Z}_{t2}$ together with convenience yields of 2008 (right panel) in Figure 4.13. The convenience yield is usually derived within a no-arbitrage or cost-of-carry model which is based on considerations on a hedging strategy consisting of holding the underlying asset of the futures contract until maturity. Assume that we hold one unit of emission rights at time t, the current spot price is S_t and the futures with maturity T are denoted with $F_{t,T}$. Obviously there is no physical storage cost for holding an emission right. Hence, assuming the existence of a convenience yield, holding the emission right until T will pay us the stochastic return

$$S_T - S_t + \mathrm{c}_{(T-t)}.$$

Hereby, $\mathrm{c}_{(T-t)}$ denotes the convenience yield for holding the emission right from t until T. Additionally for $\mathrm{c}_{(T-t)}$ we get with the following equation for the convenience yield

$$\mathrm{c}_{(T-t)} = S_t e^{r(T-t)} - F_{t,T}.$$

The convenience yield obtained from holding a commodity can be regarded as being similar to the dividend obtained from holding a company's stock. Here the necessary risk free rates were obtained using 3-month and 6-month EURIBOR rates for short-term periods and swap based zero coupon yields for the long-term interest rates up to 2012. To match the yields for different time horizons we used linear interpolation.

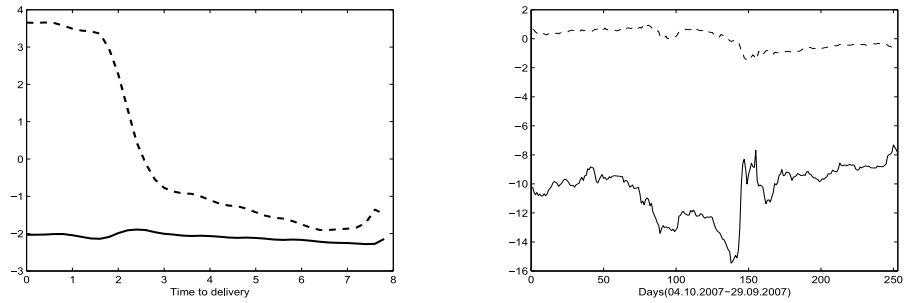

Figure 4.12: *Left panel: estimated basis functions* $\widehat{m}_1$ *(lower curve)*$\widehat{m}_2$ *(upper curve) Right panel: time series* $\widehat{Z}_{t1}$ *(solid line) and* $\widehat{Z}_{t2}$ *(dashed line).*

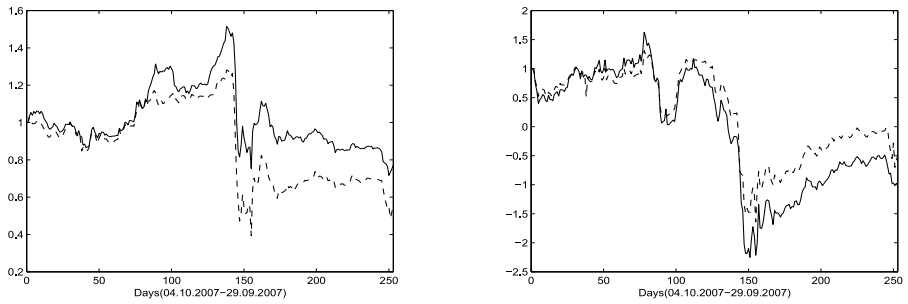

Figure 4.13: *Left panel: spot prices together with* $\widehat{Z}_{t,1}$*. Right panel: convenience yields for the year 2008 together with* $\widehat{Z}_{t,2}$*. All series are scaled to the first day for the visual purpose.*

In Figure 4.13 all the time series are rescaled by the first day of the sample in order to achieve the visual comparison of the dynamics. As it is expected the main factor, which drive the level of the term structure, is closely related to dynamics of the spot. The second factor, however, mirrors the convenience yields of the Kyoto period. This result confirms the finding of Trück et al. (2006) that convenience yields can be interpreted as the market participants expectations on allocations for the commitment trading period. When the corresponding $c_{(T-t)} < 0$ the future prices of the commitment period are

relatively high comparing to the future prices in the pilot period, which could be caused by perceived limitation of the allowances. Similarly the flattening of the term structure, which is related to convenience yields close to 0, corresponds to the lack of strong expectations on number of allowances.

The presented DSFM analysis can serve only as describing tool. Due to the highly temporary structure of future prices caused by the conversion from pilot period to the Kyoto period one may not expect good out-of-sample forecast based on this model.

4.2.3 Functional Magnetic Resonance Imaging

In this section we show how to apply the model to the fMRI data. The DSFM share the same principle with dynamic component analysis described in Thirion and Faugeras (2003). It belongs to exploratory techniques like PCA or ICA. Here we follow this spirit by using no prior knowledge and apply the model to the raw data.

As a exploratory variable X_t we take the 3 dimensional voxel's index (i_1, i_2, i_3). For the numerical tractability we reduce the initial data by taking only every second slice in each direction, which reduces our data to the series of $32 \times 32 \times 15$ cubes. We estimate the model using 405 quadratic tensor B-splines on equidistant knots. In order to identify the model we take orthogonal $\widehat{m}$ and order the factors with respect to variance. Additionally we norm $\widehat{Z}_t$ such that they have 0 mean.

For the choice of L we study like in the previous sections $RV(L)$. The results are reported in Table 4.6. In contrast to IV and CO_2 application the table does not give strong evidence on the choice of the number of factors. For our further analysis, however, we choose the model with $L = 4$.

No. Factors	$1 - RV(L)$
$L = 1$	$871.895 \cdot 10^{-3}$
$L = 2$	$871.928 \cdot 10^{-3}$
$L = 3$	$871.941 \cdot 10^{-3}$
$L = 4$	$871.951 \cdot 10^{-3}$
$L = 5$	$871.956 \cdot 10^{-3}$

Table 4.6: *Proportion of the explained variation by the models with* $L = 1, \ldots, 5$ *dynamic factors.*

The estimated functions $\widehat{m}_0 - \widehat{m}_4$ and time series $\widehat{Z}_{t1} - \widehat{Z}_{t4}$ are plotted in Figures 4.14 and 4.15 respectively. The function $\widehat{m}_0$ can be recognized as smooth version of the overall mean of the signal. Comparing this function

with the raw data in Figure 4.5 one may recover the general shape of the brain although some local patterns are smoothed out. By construction the first factor and loadings incorporate the largest variation. Note that in Figure 4.15 the time series are kept on the different scale since otherwise the shape of $\widehat{Z}_{t2} - \widehat{Z}_{t4}$ would be unrecognizable. One may see the strong positive trend in $\widehat{Z}_{t1}$ and relatively flat pattern of the $\widehat{m}_1$ spread uniformly over the brain. This effects could be typically explained by the mixture of several components like physiological pulsation, subtle head movement, machine noise etc. For the description of different artifacts, which significantly influence the fMRI signals we refer to Biswal et al. (1995). Functions $\widehat{m}_2 - \widehat{m}_4$ have a clear peak and $\widehat{Z}_{t2} - \widehat{Z}_{t4}$ show rather mild mean reverting behavior. $\widehat{Z}_{t2} - \widehat{Z}_{t3}$ oscillate around 0, while $\widehat{Z}_{t4}$ reveals more periodic type of behavior. However, it is not straightforward to give an insightful explanation, for which one requires far deeper studies.

In order to check how the recovered signal interacts with the given stimuli we plot in Figure 4.16 the response of $\widehat{Z}_t$ to the impulse, namely $\widehat{Z}_{t+s} - \widehat{Z}_s$ where s is the time of the stimuli appearance. The mean response of $\widehat{Z}_{t1}$ and $\widehat{Z}_{t3}$ shows mild similarity to the hemodynamic response, see Worsley et al. (2002), while the response of $\widehat{Z}_{t2}$ and $\widehat{Z}_{t4}$ seem to be independent from the stimuli.

The fitting procedure to the fMRI data does not use external information on the signal. From the biological perspective it could be hardly expected that pure statistical algorithm yields full insight into understanding the complex dynamics of MR images. For this purpose one has to include additional adjustment of the algorithm like the shape of hemodynamic response or consider physiologically motivated identification of the factors. It goes however beyond the scope of this illustrative example.

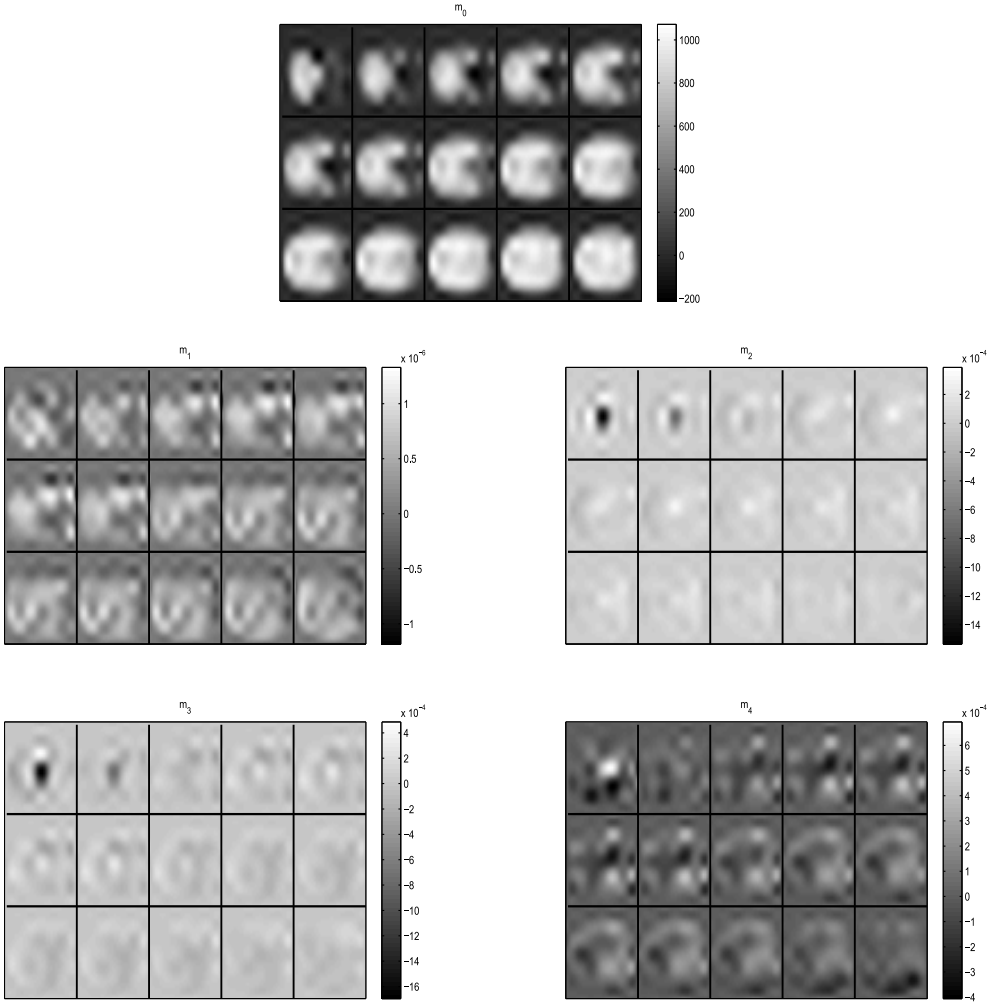

Figure 4.14: *Estimated functions on the fMRI signals:* $\widehat{m}_0$ *(upper panel),* $\widehat{m}_1$, $\widehat{m}_2$ *(middle panel),* $\widehat{m}_3$, $\widehat{m}_4$ *(bottom panel).*

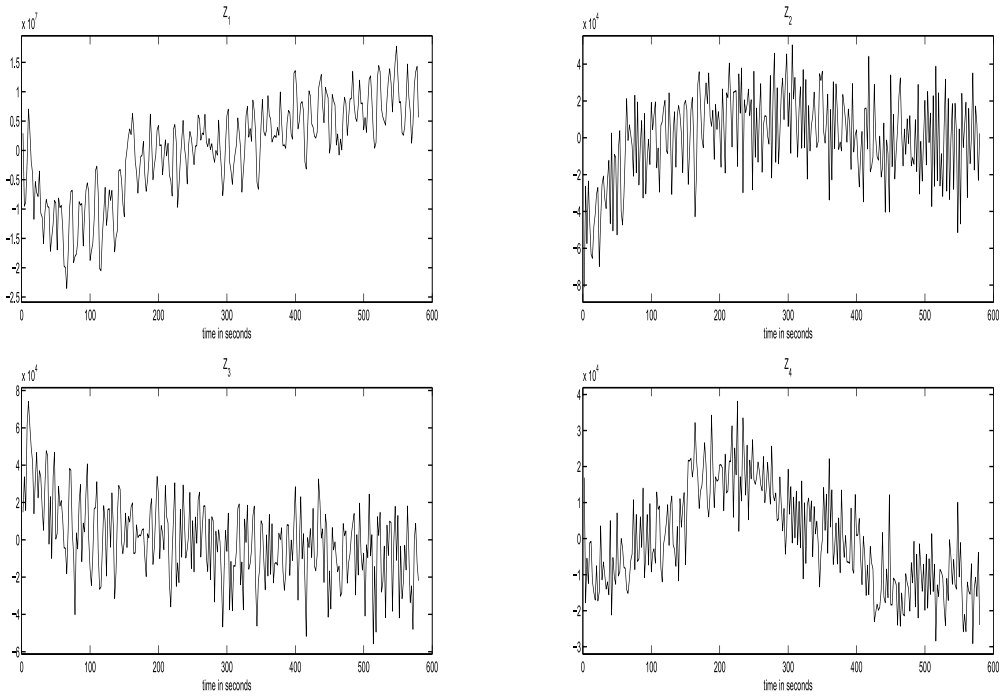

Figure 4.15: *Estimated factor time series on the fMRI signals:* $\widehat{Z}_1$ *and* $\widehat{Z}_2$ *in the upper panel,* $\widehat{Z}_3$ *and* $\widehat{Z}_4$ *in the bottom panel. The time series are displayed on the time scale in seconds although the measurements are taken every 2 seconds.*

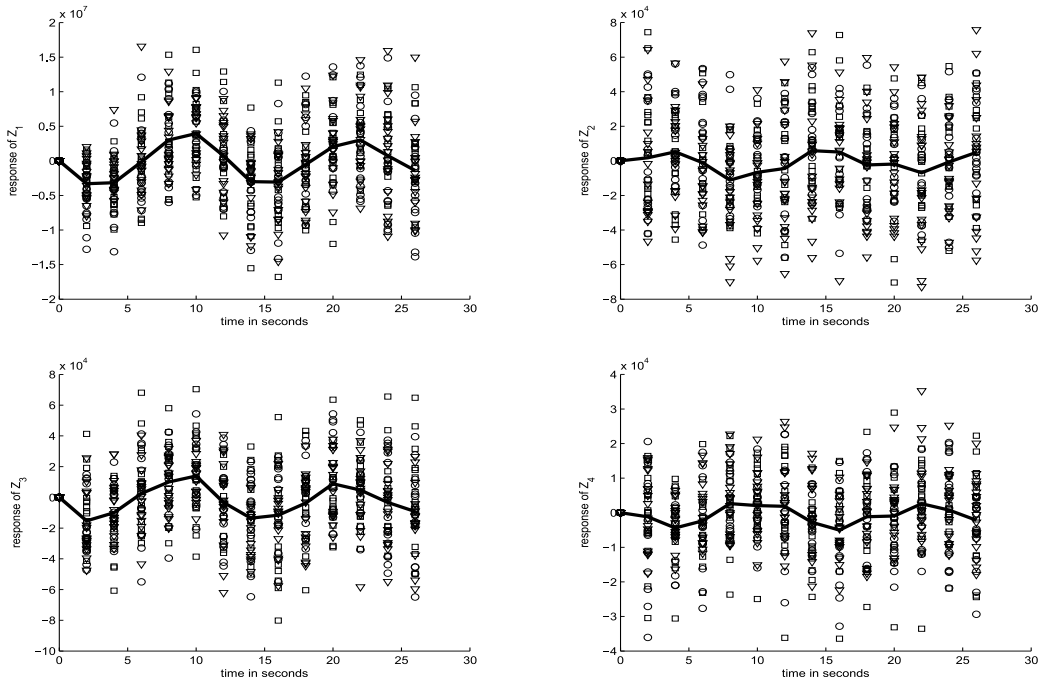

Figure 4.16: *Response of the stimuli of $\widehat{Z}_{t1} - \widehat{Z}_{t4}$. The solid line represents the mean response $\widehat{Z}_{t+s} - \widehat{Z}_s$, where s is the time of the stimuli appearance. Circles, triangles and squares denote the response of $\widehat{Z}_t$ on motorbike, bench and phone respectively.*

4.2.4 Electricity Futures

In Section 4.1.2 we presented how the DSFM can be fitted to the term structure of the CO_2 emission allowance prices. In this section we fit the model to data on the electricity futures and forward prices discussed in Section 4.1.4. Contrary to the previous sections the size of the model L is not specified but we perform the joint analysis for $L = 3, 4, 5, 6$. Additionally we compare the DSFM with one version of the fPCA approach presented in Section 2.8.2.

Figure 4.7 shows the term structure of the electricity prices on four different days. Due to the trading regulations this curve has a special piecewise constant shape. It could be observed that it is a highly dynamic object and the explanation of this dynamics is approached through the DSFM and the PCA.

The analysis is performed in the moving window framework. For the calibration of the two considered factor models we take the 500 trading days. We shift our sample by one day discarding the oldest observation and including the consecutive day. The procedure is stopped 125 days before the end of our sample since in Section 4.3 we study performance of the out-of-sample fit up to 125 days horizon. As a consequence we have 862 estimates of the two models each of the size $L = 3, 4, 5, 6$.

For the DSFM estimation we take into account the data points with maturity fixed to the middle of the delivery period. We apply the series estimator (2.59) with weights. It leads to minimizing

$$\sum_{t=1}^{T} \sum_{j=1}^{J_t} \left\{ Y_{tj} - Z_t^\top A \psi(X_{tj}) \right\}^2 w(X_{tj}). \tag{4.5}$$

Note that implementing weight does not cause any particular estimation problems since they can be easily taken inside the square difference term. Consequently the observation (X_{tj}, Y_{tj}) is weighted proportional to the length of the delivery period. For the basis functions ψ 19 cubic B-splines evaluated on the equidistant knots are used. The number of spline functions is related to the mean number of observations per day J_t and as long as the order of splines is greater than one we find the placement of the knots and selection of K of minor importance to the in-sample fit performance. For a comparison see a simulation example in Chapter 3.

The second factorial method that we study in this section is the fPCA evaluated on discrete grid of 1446 points, where each point corresponds to half day interval. Due to such a choice one can catch better the middle point of the delivery period. The discrete curves are then treated as matrix Y and the PCA is applied, see Section 2.7.3 and representation (2.36). Note that regular grid representation precludes the weighting necessity.

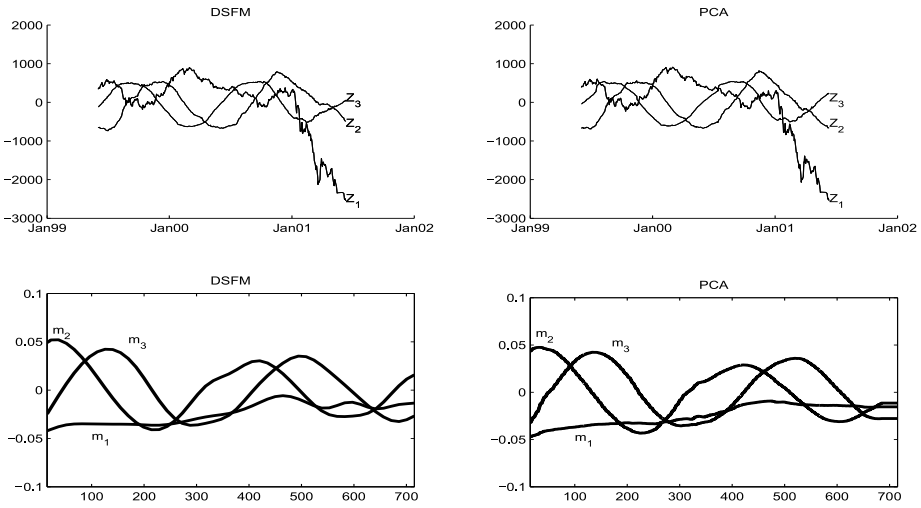

Figure 4.17: *Estimated $\widehat{Z}_{tl}$ (upper panels) and $\widehat{m}_l$ (lower panels) on the data from June 3rd, 1999 to June 11th, 2001. The models (DSFM left panels, PCA right panel) are estimated for $L = 3$. We omit presenting the $\widehat{m}_0$.*

Typical estimation result for $L = 3$ on a period from June 3rd,1999 to June 11th, 2001 is displayed in Figure 4.17. Both models exhibit similar properties. The functions $\widehat{m}_1$ are relatively flat without changing sign on the whole domain, which could be interpreted as an overall level changes. Their absolute values decreases with maturities, which is inline with the biggest variation close to expiry, observed also by Koekebakker and Ollmar (2005). The factor $\widehat{Z}_{t1}$ reflects then the trend of the entire term structure. The second and third elements of the models show the periodic behavior both in spacial loading functions and time dependent factors. The period is approximately one year and the factors can be interpreted as seasonal adjustment of the curve. Obviously in order to give the full overview of the estimates one has to present all 862 plots like Figure 4.17, which is impossible here. Naturally, there are some deviations from this pattern through the whole sample but the presented one, namely a trend factor and two seasonal factors, dominates in the data. This structure has a direct consequences in our out-of-sample setting.

For the comparison of the methods one has to consider some loss function. Among the plenty possible choices we define two types of errors

$$\epsilon_{L_1} = \sum_{t=1}^{T} \sum_{j=1}^{J_t} |Y_{tj} - Model(X_{tj})| w(X_{tj}) \tag{4.6}$$

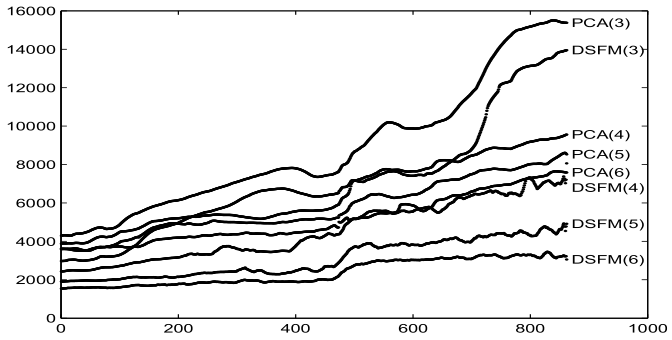

Figure 4.18: *In-sample error ϵ_{L_1} as a function of time for the DSFM and PCA for the different model size. The first sample is estimated on 500 data point from January 4th, 1999 to December 29th, 2001 the last one on the data from December 23rd, 2002 to December 30th, 2004.*

and

$$\epsilon_{L_2} = \sum_{t=1}^{T} \sum_{j=1}^{J_t} \left\{ Y_{tj} - Model(X_{tj}) \right\}^2 w(X_{tj}). \tag{4.7}$$

The term $Y_{tj} - Model(X_{tj})$ reflects how well the model replicates the observed data points while weight $w(X_{t,j})$ enhance the importance of the data point. Quadratic error ϵ_{L_2} is in fact minimized by the DSFM. We have not found any qualitative differences between (4.6) and (4.7) and for our further analysis we consider more robust error ϵ_{L_1}. Note that the resulting term structure from the factor models is a smooth function and we impose the piecewise constant shape by evaluating models only in the middle points of the delivery period.

The in-sample error as function of time is presented in Figure 4.18. Clearly the models with more factors give better fit although for models estimated in the beginning of our sample it is less pronounced than for the models from the end of the sample. For all cases one observes the upward trend which means that the models with same number of factors fit much better the data of years 1999-2001 than years 2001-2004. In particular it is visible for $L = 3$.

The factor models estimated with the DSFM yield superior fit than the models estimated with the PCA. It is particularly true for the models with the same number of factors, however, the DSFM with $L = 4$ gives mostly better fit than PCA with $L = 6$. It could be also seen in Figure 4.19, where we show the quantiles and boxplots across the whole sample.

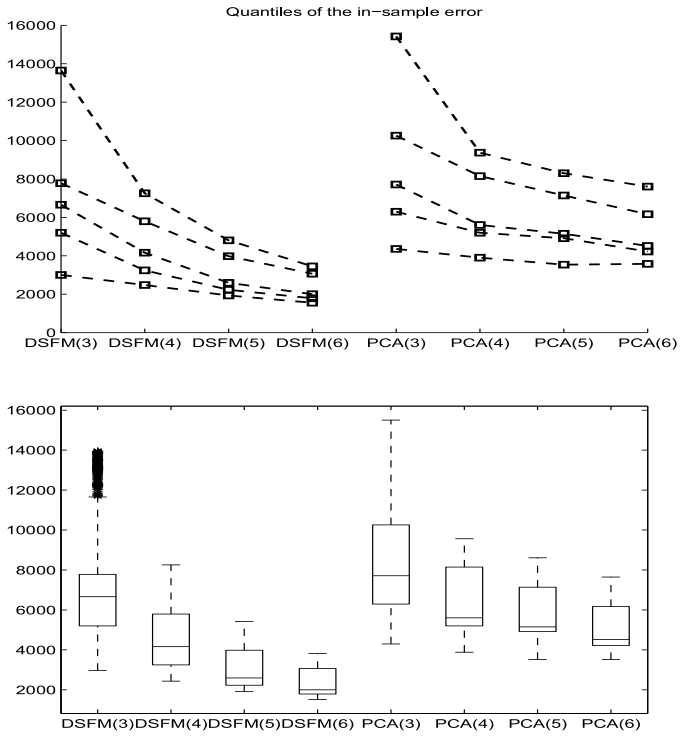

Figure 4.19: *Upper panel: 2.5, 25, 50, 97.5, quantiles of the in-sample error* ϵ_{L_1} *over all samples for DSFM (left) and PCA (right).*

4.3 Forecasting

In Section 4.2.4 we show the DSFM fit to the electricity futures data in a moving window framework. Here our aim is to extend this study to the out-of-sample fit. We use the estimated model to predict the whole curve in 1 day, 5 days, 25 days, and 125 days horizons. The results are compared to the results obtained by the PCA. This study has the same motivation as the analysis on the artificial data from Section 3.5.1. Here the real data are used and in order to obtain significant number of observations a moving window has to be applied.

We evaluate several types of predictions, which seem to be reasonable for the discussed issue. They are based on the inclusion of the estimation error, structure of the estimated factors and the randomness of the factors. The settings are described in Section 4.3.1. The results are summarized mainly in graphical form in Section 4.3.2.

4.3.1 Setting

In this section we adopt the functional notation in order to give more intuitive definitions of our prediction methods. Let x denotes the maturity of the futures contract and the term structure in time point t be described by a function $Y_t(x)$. For the electricity futures this function is a piecewise constant curve, see Figure 4.7. Let $Y^*_{T+h}(x)$ be the forecast in h days, where T is the index of the last observed curve. The out-of-sample error may be defined similarly as in Section 4.2.4. Here we consider error (4.6) as well.

For the first approach to forecasting an electricity futures term structure, we follow the famous sentence from Bachelier: *The best prediction for the value tomorrow is the value today.* We call it random walk (RW) and it can simply be defined as

$$Y^*_{T+h}(x) = Y_T(x). \tag{4.8}$$

It assumes that there is no structure in the dynamics and the best prediction for the future curves is the observed curve itself. This approach has perfect in-sample fit for the time point T and in case of small variation it is likely to expect that it would yield a reasonable forecast in the short time horizon.

Another idea is to update the $Y_T(x)$ by the trend hidden in the estimated factors $\widehat{Z}_t$. Figure 4.17 shows that $\widehat{Z}_{t2}$ and $\widehat{Z}_{t3}$ reflect periodic behavior. This trend can be easily removed by least square fit of the sin function of the form $\widehat{g}_l(t) = a_l \sin(b_l t + c_l)$. Then the forecast has the form

$$Y^*_{T+h}(x) = Y_T(x) + \sum_{l=2}^{3} \widehat{m}_l(x) \left\{ \widehat{g}_l(T+h) - \widehat{g}_l(T) \right\}. \tag{4.9}$$

We have not found the seasonal behavior in the other factors so the summation is only over the second and third one. This setting we name (STr1) and (PTr1) for the DSFM and the PCA respectively.

Previously one updates the observed curve by the sinusoidal trend. However, it is also possible to update the curve resulted from the estimated model. This setting called (STr2) and (PTr2) has the form

$$\begin{aligned} Y^*_{T+h}(x) &= \sum_{l=0}^{L} \widehat{Z}_{Tl}\widehat{m}_l(x) + \sum_{l=2}^{3} \widehat{m}_l(x)\left\{\widehat{g}_l(T+h) - \widehat{g}_l(T)\right\} \\ &= \sum_{l=0}^{L} \widetilde{Z}_{Tl}\widehat{m}_l(x) + \sum_{l=2}^{3} \widehat{m}_l(x)\widehat{g}_l(T+h), \end{aligned} \tag{4.10}$$

where $\widetilde{Z}_{tl} = \widehat{Z}_{tl} - \widehat{g}_l(t)$ with natural choice $g_l(t) = 0$ if $l \neq 2, 3$. The only difference to (4.9) is discarding the estimation error of the last day in estimation sample.

A natural extension of (4.10) is considering an additional forecast for the time series $\widetilde{Z}_{tl}$ for $l > 0$. This setting (STS, PTS) can be presented as

$$Y^*_{T+h}(x) = \sum_{l=0}^{L} \widetilde{Z}^*_{T+h,l}\widehat{m}_l(x) + \sum_{l=2}^{3} \widehat{m}_l(x)\widehat{g}_l(T+h). \tag{4.11}$$

For the $\widetilde{Z}^*_{T+h,l}$ we take AR(2) forecast. We studied also other types of time series models but the differences are negligible so for the clearness of the presentation we keep AR(2) since it yields slightly better results.

4.3.2 Results

The comparison across different methods is summarized in Figures 4.20-4.23. We mainly focus on the quantiles of the empirical distribution and for this reason we have chosen boxplots for our illustration. Obviously lower boxplots correspond to smaller forecasting errors. We keep $L = 4$ but the results for other number of factors do not differ significantly.

For the short time horizon (1 day) the methods which take into consideration $Y_T(x)$ outperform the model based methods. It practically means that *the value tomorrow* does not change much and the estimation error of the models preclude the successful short time prediction. One cannot observe significant difference among RW, STr1 and PTr1. The DSFM based approaches STr2 and STS perform slightly better than PTr2 and PTS. This could be explained with the better in-sample fit of the semiparametric method. Very similar pattern is visible in Figure 4.21, where 5 days prediction is presented.

Here, however, PTr2 and PTS have similar forecasting error distribution as STr2 and STS.

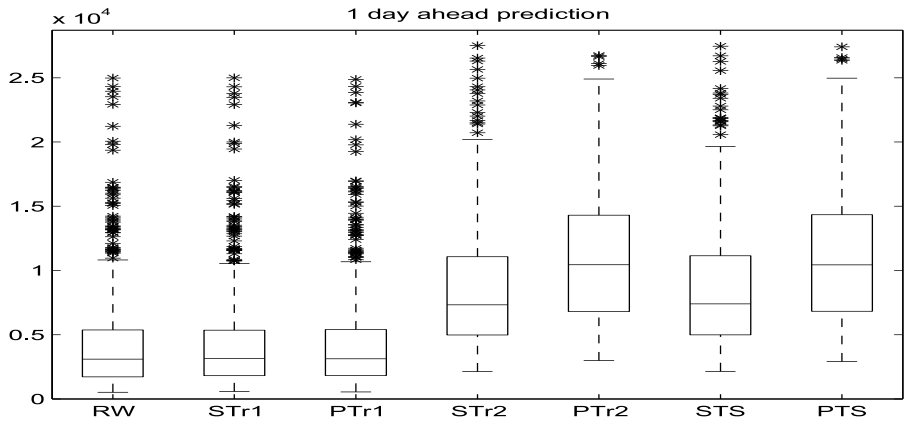

Figure 4.20: *Out-of-sample errors for horizon $h = 1$ and different prediction methods.*

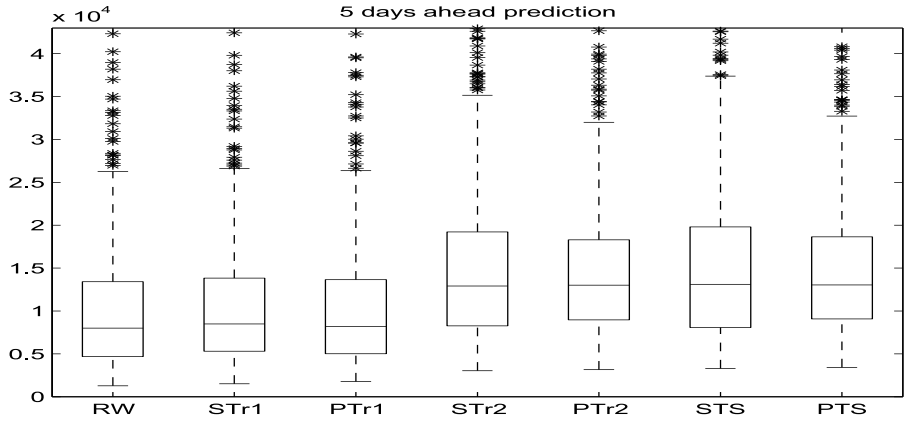

Figure 4.21: *Out-of-sample error for horizon $h = 5$ and different prediction methods.*

For the middle time prediction (25 days) we refer to Figure 4.22. RW, STr1, PTr1 performs still similarly but PTr2 and PTS yield comparable results. The biggest errors in terms of quartiles are for STr2 and STS. The long time horizon (125 days) is presented in Figure 4.23. The RW is no longer a reasonable predictor. In terms of quartiles and median STr1, STr2, STS are better than RW. However, it is nothing unexpected since there is

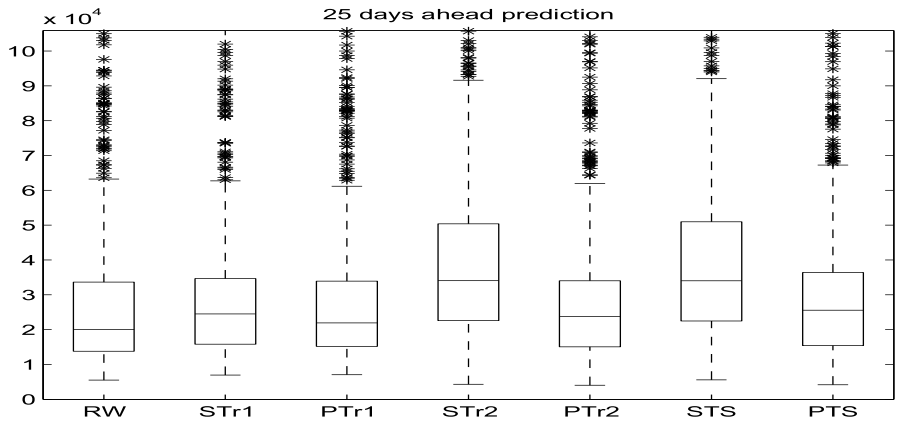

Figure 4.22: *Out-of-sample error for horizon $h = 25$ and different prediction methods.*

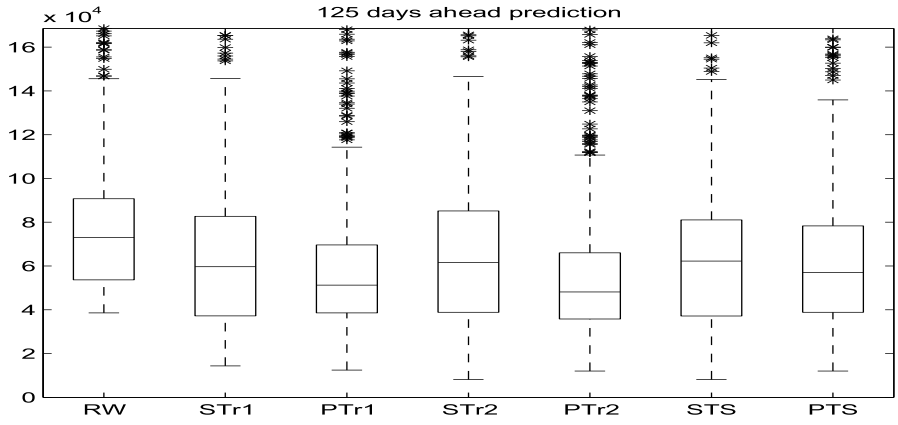

Figure 4.23: *Out-of-sample error for horizon $h = 125$ and different prediction methods.*

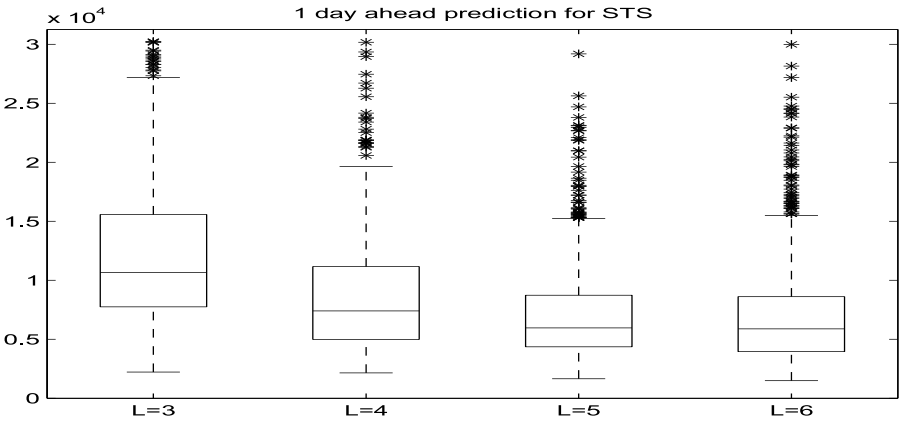

Figure 4.24: *Out-of-sample errors for different number of factors and time horizon 1 day. STS method is considered.*

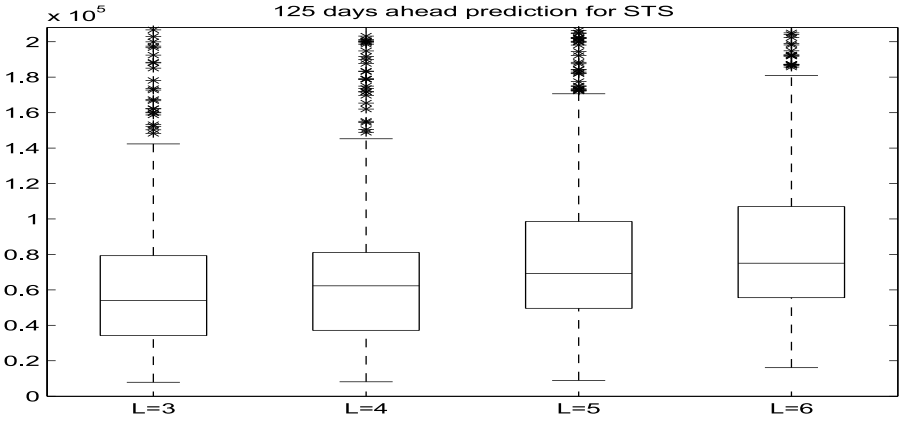

Figure 4.25: *Out-of-sample errors for different number of factors and time horizon 125 day. STS method is considered.*

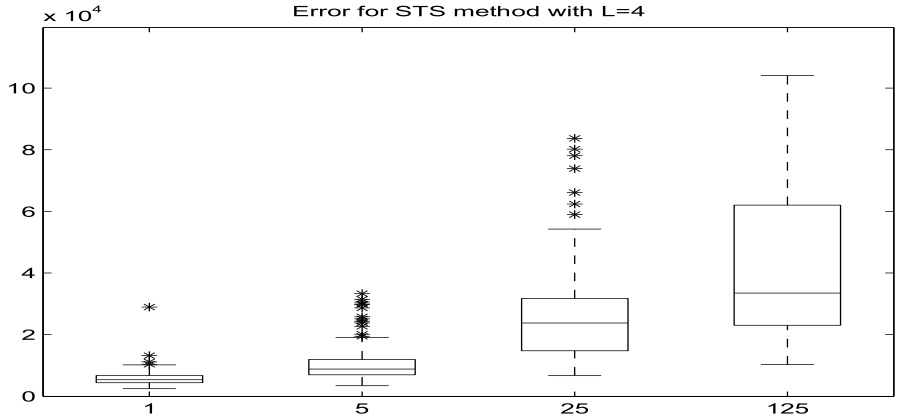

Figure 4.26: *Out-of-sample error with different prediction horizons for STS method with $L = 4$ factors.*

sinusoidal behavior of the whole curve and 125 days corresponds to the half of the period. Therefore RW generates just inverted curves. The PCA based methods have better performance than their DSFM counterparts. Besides PTr2 seem to give best results.

The presented analysis shows that in short and medium time horizons the methods with incorporated estimation error performs relatively better to the methods based only on the statistical model. It means that the reversion to the mean is not so fast and the models do not bring significant improvements in the short time period. In the long run the inclusion of trends improve the random walk however no significant improvement comes from the time series analysis. The PCA based methods perform better than the DSFM methods. It may be explained by the overfitting of the latter model. The model fits too good to the estimation data set and therefore has inferior out-of-sample fit. To overcome this problem one may impose some roughness penalty and simultaneously deteriorate the in-sample fit.

To give the full picture we present additionally in Figure 4.24 and Figure 4.25 how the STS setting performs for different number of factors. The 1 day and 125 days horizons are presented respectively. For the short time forecast inclusion additional factor improve the prediction. This comes from the better in-sample fit. For the long time forecast, however, the prediction with six factor model is worse than with $L = 3$ or $L = 4$. This observation confirms that better in-sample fit does not necessarily generate better forecasting. As a last result we show in Figure 4.26 how the error depends on the time horizon. As it could be expected the longer horizon the bigger the forecasting errors. Their variation increases as well accordingly.

4.4 Hedging

The aim of this section is to perform an empirical study on dynamic hedging of reverse barrier options in the presence of the IVS dynamics. The reverse barrier options are the ones having their barrier in-the-money, i.e. breaching the barrier results in lost of the full intrinsic value. For various reasons discussed below we assume a local volatility (LV) model for the underlying. A price of barrier options in the LV model strongly depends on a shape of an entire IVS. Since IVS is a highly volatile object, as indicated in Section 4.1.1 and 4.2.1, the barrier options are exposed to nontrivial volatility risk. We extract the key risk factors resulting from the implied volatility surface fluctuations by the DSFM and on this basis define a practical hedging procedure in local volatility framework.

The hedging performance is evaluated by empirical investigation on the DAX index options. We establish an artificial portfolio of reverse barrier option and hedge it on a daily basis with plain vanillas and underlying taking into account the true market conditions. Our methodology follows the one presented in Engelmann et al. (2006) and their hedging results we treat as a benchmark.

4.4.1 Motivation

In equity derivative markets barrier options became appealing alternative instruments for investors searching for nonconventional payoff profile. In order to respond for this demand banks structure investment products in form of bonus certificates, which contain these options. Such a business activity results in holding large short position in banks' trading books. As a consequence efficient hedging strategies have to be implemented. This is a challenging task in particular for reverse barrier options, which knock out deep in-the-money loosing the maximum possible intrinsic value.

The dynamic hedging performance of the LV model for the reverse barrier options is studied in Engelmann et al. (2006). Their hedge strategy constructs delta ($\partial/\partial S$), vega ($\partial/\partial\sigma$) and vanna ($\partial^2/\partial\sigma\partial S$) neutral portfolios. The sensitivities vega and vanna are obtained by a parallel shifts of the IVS and computing the difference quotient. While an up-and-down shift of the IVS may be the most important factor, the parallel-shift-vega leaves the slope and term structure risks, which the exotic option is exposed to, completely unhedged. Depending on the specific payoff profile of the options, these risks, however, can be of substantial size. For instance, for down-and-out puts, the probability of hitting the barrier is very much determined by the slope of the smile. In this case, it is therefore desirable to hedge the slope

risks of the IVS. In our hedging strategy we define new sensitivity measure with respect to the most common IVS movements motivated by the DSFM, namely $(\partial/\partial Z_{t1})$, $(\partial/\partial Z_{t2})$ and build portfolio neutral to this greeks.

We have to remark here that there exist a vast literature on the static hedging of barrier options, see Carr et al. (1998), Derman et al. (1995), Carr and Chou (1997), Andersen et al. (2002), Nalholm and Poulsen (2006a), Nalholm and Poulsen (2006b), Tompkins (2002). In static hedge one sets up portfolio of plain vanillas which replicate as close as possible the payoff of the barrier option. The hedge is unwind in case of knock-out or in expiry and besides no other restructuring is committed. In fact, Engelmann et al. (2007) show that some static hedge outperform the dynamic hedge. However, the static hedge cannot always be practically implemented due to the insufficient market depth. Moreover the banks' traders perform on the constant time basis and update they hedge ratios relatively frequent. Therefore we omit here the static hedging problematic and focus on studying the dynamic hedge.

4.4.2 Models

In this section we present the LV model that is applied for pricing and hedging the barrier options. We also discuss factor hedging related issues resulting from the DSFM fitting for IVS dynamics, which are later utilized in the hedging framework.

In the LV model the risk neutral price of the underlying is governed by the stochastic differential equation

$$dS_t = (r_t - q_t)\, S_t dt + \sigma(S_t, t) S_t dW_t, \tag{4.12}$$

where W_t is a Wiener process, r_t denotes the instantaneous interest rate and q_t is a dividend yield. From now on for the simplicity of notation we assume $q_t = 0$. $\sigma(S_t, t)$ is the local volatility function which depends on the underlying price and time. This surface has a unique representation if the arbitrage-free call options are given for all strikes and maturities, see Dupire (1994). The model (4.12) yields then the correct vanilla prices. It can be shown that

$$\sigma^2(S_t,t) = \left. \frac{2\frac{\partial\widehat{\sigma}(\mathrm{K},T)}{\partial T} + \frac{\widehat{\sigma}(\mathrm{K},T)}{T} + 2\mathrm{K}r_T\frac{\partial\widehat{\sigma}(\mathrm{K},T)}{\partial \mathrm{K}}}{\mathrm{K}^2\left\{\frac{\partial^2\widehat{\sigma}(\mathrm{K},T)}{\partial \mathrm{K}^2} - d_1\sqrt{T}\left(\frac{\partial\widehat{\sigma}(\mathrm{K},T)}{\partial \mathrm{K}}\right)^2 + \frac{1}{\widehat{\sigma}(\mathrm{K},T)}\left(\frac{1}{\mathrm{K}\sqrt{T}} + d_1\frac{\partial\widehat{\sigma}(\mathrm{K},T)}{\partial \mathrm{K}}\right)^2\right\}} \right|_{\mathrm{K}=S_t, T=t} \tag{4.13}$$

where $d_1 = \frac{\log(S_0/\mathrm{K}) + \int_0^T r_s ds + 0.5\widehat{\sigma}^2(\mathrm{K},T)T}{\widehat{\sigma}(\mathrm{K},T)\sqrt{T}}$ and $\widehat{\sigma}(\mathrm{K},T)$ is implied volatility taken

in strike K and maturity T. The formula (4.13) stresses the correspondence between local and implied volatility surfaces.

The LV model aroused great attention in the financial literature and was simultaneously criticized on several issues. Ayache et al. (2004) point out the lack of the physical explanation of the smile phenomenon contrary to stochastic volatility or jump diffusion models. Besides, it predicts too shallow smiles at future dates, which makes it inappropriate for pricing options with a long-dated forward-start period. Furthermore, Hagan et al. (2002) show that it implies unrealistic smile dynamic. The positive movement of the underlying moves the IVS to the left, which is opposite to the typical market behavior, in which implied volatility skew moves in the same direction as the underlying. This behavior may lead to unstable delta and vega hedges.

Despite the critique we believe the LV model serves well for the purpose of our empirical study. Due to its nonparametric character there is no calibration bias and the prices of plain vanilla options used for hedging exotics are fully consistent with market prices. Since we consider only one year barrier options the flatter LV surface at long-dated horizons seems to be of minor importance. The problem with incorrect smile dynamics is often solved in practise by enforcing the desired IVS dynamics. Instead of calculating model-consistent LV greeks, one fixes the IVS in strikes (sticky-strike) or in moneyness (sticky-moneyness) and recalibrate the LV surface under the spot movements. In fact Engelmann et al. (2006) find that the empirical performance for the dynamic hedging of reverse barrier options is negligible under different stickiness assumptions, unless the vega hedge is implemented. Nevertheless, they claim that sticky-strike approach performs best and we also adopt it in this thesis.

However, the most prominent advantage of the LV model is its numerical tractability. The price of the barrier option denoted as V with barrier B and expiring at T is obtained through numerical solution of the partial differential equation

$$r_t V(S,t) = \frac{\partial V(S,t)}{\partial t} + \frac{1}{2}\sigma^2(S,t)S^2\frac{\partial V(S,t)}{\partial S^2} + r_t S\frac{\partial V(S,t)}{\partial S} \tag{4.14}$$

with additional boundary conditions i.e. $V(B,t) = 0$ for $t < T$ and $V(S,T)$ equal to the payoff at expiry. For calibration of the model a number of methods are available, see Bouchouev and Isakov (1999) for comprehensive review. For example one may directly apply the formula (4.13). Here we adopt more stable approach of Andersen and Brotherton-Ratcliffe (1997) which determines r and σ so that forwards, zero coupon bonds and plain vanilla options are priced correctly on each grid point. Then the finite difference method gives barrier options and sensitivities very efficiently. It is of main impor-

tance since the sticky-strike delta and the whole bunch of the volatility greeks requires frequent recalibration of the model under different IVS scenarios.

For statistically motivated IVS dynamics we employ the DSFM. We chose the kernel estimation method (2.48). In order to illustrate the meaning of the DSFM for the IVS dynamics we present in Figure 4.27 the result of the estimation on the DAX option data from January 3rd, 2000 till June 30th, 2004. The figure presents the estimated $\widehat{Z}_{tl}$ time series in the upper panel and estimates of the basis functions in the middle panel. To save the space we do not present the invariant function $\widehat{m}_0$, which has no effect on the dynamics of the IVS but has to be included to set the correct level of the surface. The function $\widehat{m}_1$ is relatively flat, although not completely flat, and corresponds to the most common shocks. Changes in $\widehat{Z}_{t1}$ results in up-and-down type of movements of the whole surface, although the discrepancies from the complete flat pattern gives different weight in each maturity/moneyness location. This effect is illustrated in the lower panel, where we plot one particular smile with different values of $\widehat{Z}_{t1}$. The second dynamic factor could be interpreted as a tilting of the smile. It can be seen in the shape of $\widehat{m}_2$ and the influence plot in the lower panel. The variation in $\widehat{Z}_{t2}$ results rather in changing the slope of the smile by making it more stepper or flatter and keeping relatively same level.

4.4.3 Hedging Framework

The dynamic hedging of the asset V, in our case reverse barrier options, utilize the frequent updates of the hedge portfolio. For this aim one calculates the sensitivity measures and tries to create a portfolio neutral to the market changes. Apart from standard delta hedging the successful strategy requires higher order greeks, in particular volatility greeks, as pointed out by Ederington and Guan (2007).

In LV model the different dynamic hedging strategies of reverse barrier options are studied by Engelmann et al. (2006). They consider delta, delta-vega and delta-vega-vanna hedges. A one unit of knock-out option is hedged with underlying and set of plain vanilla options. Let the value of the barrier option be denoted by V and let HP_1 and HP_2 be portfolios of plain vanillas. Then the corresponding hedge ratios are given by solving

$$\begin{pmatrix} 1 & \frac{\partial HP_1}{\partial S} & \frac{\partial HP_2}{\partial S} \\ 0 & \frac{\partial HP_1}{\partial \widehat{\sigma}} & \frac{\partial HP_2}{\partial \widehat{\sigma}} \\ 0 & \frac{\partial^2 HP_1}{\partial \widehat{\sigma} \partial S} & \frac{\partial^2 HP_2}{\partial \widehat{\sigma} \partial S} \end{pmatrix} . \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial V}{\partial S} \\ \frac{\partial V}{\partial \widehat{\sigma}} \\ \frac{\partial^2 V}{\partial \widehat{\sigma} \partial S} \end{pmatrix} . \tag{4.15}$$

The setting (4.15) reflects the full delta-vega-vanna hedge. Constraining

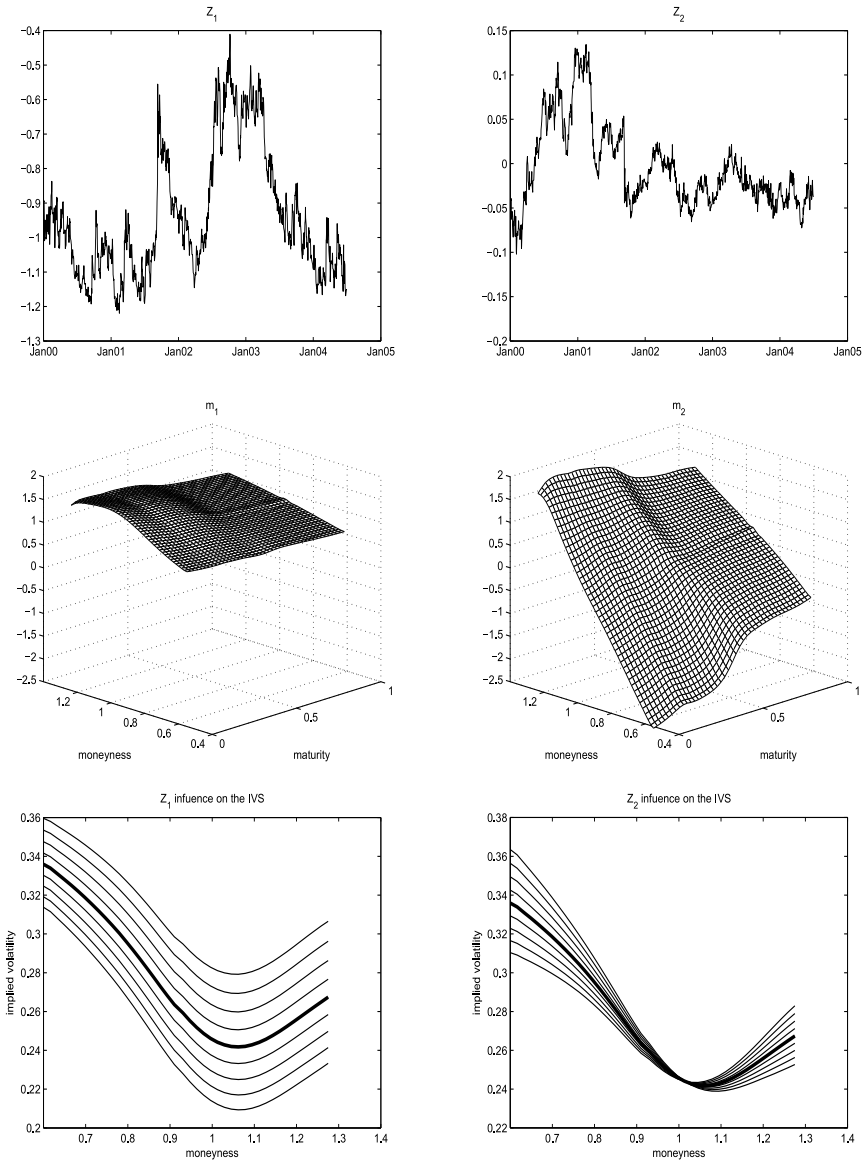

Figure 4.27: *Estimates of the DSFM on IVS from January 3rd, 2000 till June 30th, 2004. for $L = 2$. Upper panel: estimated time series $\widehat{Z}_1$ and $\widehat{Z}_2$. Middle panel: estimates of $\widehat{m}_1$ and $\widehat{m}_2$. Lower panel: impact of $\widehat{Z}_1$ and $\widehat{Z}_2$ changes on one particular smile observed on 20010103 with maturity 1 month. Shocks in $\widehat{Z}_1$ would cause up-and-down movements while shocks in $\widehat{Z}_2$ tilts the smile around at-the-money point.*

$a_2 = 0$ reduces (4.15) to delta-vega hedge and $a_1 = a_2 = 0$ to the simple delta hedge. It is desirable that HP_1 and HP_2 have large exposure to the risk factors. For the choice of HP_1 one may take at-the-money plain vanillas and for HP_2 risk reversal, which is a combination of long out-of-the-money call and short out-of-the-money put (or vice versa).

In order to approximate sensitivity measures one reprices the option under different inputs and compute the greeks by the finite difference. We stress that based on the results of Engelmann et al. (2006) we keep the sticky strike version, i.e. IVS remains constant in strikes. The vega and vanna are computed with parallel shift of the IVS. To be more specific

$$\frac{\partial V}{\partial S} \overset{\text{def}}{\approx} \frac{V\left(S + \Delta S, \widehat{\sigma}\right) - V\left(S - \Delta S, \widehat{\sigma}\right)}{2\Delta S}, \tag{4.16}$$

$$\frac{\partial V}{\partial \widehat{\sigma}} \overset{\text{def}}{\approx} \frac{V\left(S, \widehat{\sigma} + \Delta\widehat{\sigma}\right) - V\left(S, \widehat{\sigma} - \Delta\widehat{\sigma}\right)}{2\Delta\widehat{\sigma}}, \tag{4.17}$$

$$\begin{aligned}\frac{\partial^2 V}{\partial S \partial \widehat{\sigma}} \overset{\text{def}}{\approx} \Big\{ & V\left(S + \Delta S, \widehat{\sigma} + \Delta\widehat{\sigma}\right) - V\left(S + \Delta S, \widehat{\sigma}\right) \\ & - V\left(S - \Delta S, \widehat{\sigma} + \Delta\widehat{\sigma}\right) + V\left(S - \Delta S, \widehat{\sigma}\right) \Big\} / 2\Delta S \Delta\widehat{\sigma}.\end{aligned} \tag{4.18}$$

With the small abuse of notation $V(S, \widehat{\sigma})$ denotes here the price obtained with spot S and IVS $\widehat{\sigma}$, where for simplicity we omit its arguments. $\widehat{\sigma} + \Delta\widehat{\sigma}$ means the parallel shift of the whole surface, which is then incorporated in the pricing engine. Note that using sticky-strike approach the plain vanillas' greeks are identical to their Black-Scholes counterparts.

Although parallel shifts are the most prominent movements of the IVS other types of surface variations may considerably influence the prices of the barrier options. In particular the higher slope leads to smaller price of the in-the-money down-and-out put. Consider an artificial example of two one year knock-out put options with strike 110, barrier 80 and the current spot level 100. The first option is priced with IVS taken from January 3rd, 2000 and the second one from January 2nd, 2001. The prices of these options are respectively 1.91 and 2.37, which is a 25% difference. From the upper panel of Figure 4.27 one may recover that the level related factor has the similar value on this days, while slope factor differs significantly. Based on this result one may claim that the price discrepancy comes mainly from the slope effect. This exposure is not directly hedged in the approach described above. Therefore we propose to reestablish the procedure by hedging the most common vola shocks extracted by the DSFM.

The hedging framework remains similar to (4.15). Here we define new sensitivity measures with respect to the variation of the (log)-IVS, which we

call ζ-greeks. In particular based on the results discussed in Section 4.4.2 ζ_1-greek ($\partial/\partial Z_{t1}$) would reflect slightly corrected up-and-down shifts, while ζ_2-greek ($\partial/\partial Z_{t2}$) would correspond to slope effect. Then similarly to (4.15) one obtains the hedge ratios by

$$\begin{pmatrix} 1 & \frac{\partial HP_1}{\partial S} & \frac{\partial HP_2}{\partial S} \\ 0 & \frac{\partial HP_1}{\partial Z_{t1}} & \frac{\partial HP_2}{\partial Z_{t1}} \\ 0 & \frac{\partial HP_1}{\partial Z_{t2}} & \frac{\partial HP_2}{\partial Z_{t2}} \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial V}{\partial S} \\ \frac{\partial V}{\partial Z_{t1}} \\ \frac{\partial V}{\partial Z_{t2}} \end{pmatrix}. \tag{4.19}$$

The full setting we call $\zeta_1\zeta_2$ hedge and reduced one with $a_2 = 0$ ζ_1 hedge. As in traditional hedging we set at-the-money plain vanilla for HP_1 since it has the most significant sensitivity to the IVS level. For HP_2 we take risk reversals because it primarily responds to the changes of the IVS wings and by selecting the appropriate strikes it can be even set up in a vega-neutral (ζ_1-neutral) way.

The numerical approximation of the ζ-greeks can be calculated via a difference quotient. The DSFM yields the dynamics and the estimates of the IVS on each particular day, however, they are subjected to some estimation error, which may strongly influence the true price of the barrier option. Therefore in order to avoid the mispricing we do not shift by an infinitesimal weight the IVS estimates but the true surface $\widehat{\sigma}$. Thus the definition of the ζ-greek approximation is given by

$$\frac{\partial V}{\partial Z_{tl}} \overset{\text{def}}{\approx} \frac{V\left(S, \widehat{\sigma}\exp(\Delta Z_{tl} m_l)\right) - V\left(S, \widehat{\sigma}\exp(-\Delta Z_{tl} m_l)\right)}{2\Delta Z_{tl}}. \tag{4.20}$$

In practical implementation of (4.20) one faces couple of difficulties, which need to be cautiously considered. First, the size of the ΔZ_{tl} has to be optimized. Either too small or too big can distort the meaning of these greeks. Moreover, it cannot be unique for all Z_t-s and all days, since the size of the shift depends on the basis functions $\widehat{m}_l$ and the IVS on a particular day. Therefore, we propose to set the ΔZ_{tl} so that the mean upward (downward) shift is a one vola-point. It is inline with the standard vega shifts from (4.17)-(4.18) where the IVS is shifted uniformly also by one vola-point. Another challenge is an accurate calculation of barrier prices. The pricing engine takes as an input the whole IVS and adjust the grid according to the input surface. However when the two surfaces are very similar, like upward and downward shifted surfaces, the discretization error can be in the same magnitude as the price differences. This problem results in a instability of the greeks, which for example do not form a smooth function of the spot price. To overcome this problem we set the constant grid in the pricing algorithm for calculating

the ζ-greeks. The grid is fixed for the $\widehat{\sigma}$ and both $V\left(S, \widehat{\sigma}\exp(\Delta Z_{tl} m_l)\right)$ and $V\left(S, \widehat{\sigma}\exp(-\Delta Z_{tl} m_l)\right)$ are calculated on this constant grid. Furthermore, although the input IVS $\widehat{\sigma}$ is supposed to be the arbitrage free the shifted surfaces do not necessarily posses this property. Therefore, we suggest checking additionally the arbitrage conditions before calculating the ζ-greeks. For this aim we apply the algorithm of Fengler (2005a), which impose the suitable shape constraints in the option price function utilizing natural spline smoothing.

4.4.4 Experimental Design

In our empirical study we assume no transaction costs, no restriction on short selling and the possibility of trading each asset to the arbitrary size. Each security is priced using LV model calibrated to the market data from particular day. We perform the hedging strategies described in Section 4.4.3, which means we focus only on the volatility and spot risks, leaving other risks like rho risk unhedged.

In the first step of our experiment we estimate the DSFM, which is necessary for the calculation of ζ-greeks. Since in the hedging procedure only 2 main factors are included we keep $L = 2$. For the kernel function we take product quartic kernel, where $k_h(u) = 15/16(1-u^2)^2$ for $|u| < 1$ and 0 otherwise. For the data motivated bandwidths choice we refer to Section 2.9.3. However, we do not find significant differences in the estimation results unless the bandwidths are unreasonably large, which induce large bias, or extremely small which may preclude the successful estimation.

We start one long position in yearly reverse barrier options for each day until one year before the last observation. We do this because we want to evaluate all initiated hedges at market prices in the sample. Our arbitrary choice of the exotics are up-and-out call with strike at 80% of the spot and barrier at 140% as well as down-and-out put with strike at 80% and barrier at 110%. These specifications corresponds to typically traded contracts.

Based on the calibrated LV model ζ-greeks, delta, vega and vanna are calculated and the hedging strategies described in Section 4.4.3 are build accordingly. For our study we concentrate on vega, vanna, ζ_1 and $\zeta_1\zeta_2$ strategies since as indicated Engelmann et al. (2006) the pure delta hedge is inferior to the volatility based hedges. In HP_1 we use at-the-money puts for call barriers and at-the-money calls for put barriers. The risk reversal are structured by taking 80% and 120% strikes of the current spot. The hedging is financed by proper adjustment of the cash account.

The still alive positions are updated on the daily basis. This choice is motivated by the results of Engelmann et al. (2006), who do not find significant

differences in the other re-balancing frequencies. If the barrier is breached or the barrier option expire we unwind the hedge and record the hedging error. The hedge ration a_0, a_1, a_2 are adjusted, which has a obvious direct influence on the cash account.

The recorded hedge errors is given in the money values. All positions are traded at market prices and in case of knock-out event the hedge error pays interest until expiry in order to make results comparable. The cash account is interest-bearing or refinanced at the riskless short rate of the particular trading day. Altogether we have 885 hedge errors for two types barrier options with 4 different hedging strategies for each of them.

4.4.5 Results

For evaluation of the hedge performance we use a pool of 885 hedge errors. In order to make them comparable we normalize by the spot price at the time when the hedge is initiated. This normalization is common in practice and remove from the hedge errors the dependence on the underlying's level. Another normalizing factor could be the option price itself but since our risk reversals and barriers are defined as percentage of the underlying, measuring errors in relation to the spot seems to be more intuitive.

The aim of the successful hedging is possibly large reduction of the portfolio variation. In the ideal case the hedge portfolio should have zero variance and zero mean, but from the obvious reasons it may not be realized in practice. Our aim is to give a comparative analysis on the generated hedge error samples to check how the volatility factors enhance the hedging performance. We use traditional descriptive statistics (range, mean, median etc.) to assess the location and dispersion of the errors, and superior method would keep these quantities close to zero in absolute terms.

The empirical results are summarized in Tables 4.8 and 4.9 for calls and puts respectively. We present minimum, maximum, mean, median, standard deviation, absolute deviation around the median (denoted as madev.) skewness and kurtosis. The final result is given in rows marked with 0. The center is located around zero, with mean slightly below zero for the calls and slightly above for the puts. In terms of mean and median the $\zeta_1\zeta_2$ hedge preforms best for the call options but we may not have the same conclusion for the puts. Note that for the both cases the vanna hedge has the smallest minimum and maximum. Note as well that the maximum is significantly bigger in all cases, which indicates the asymmetry in the error distribution. This could also be seen in the significant skewness values.

For evaluating the variation of the hedge errors we focus the attention to std. and madev. statistics. As before there is no clear ranking among the

methods although the ones with higher order greeks that incorporate hedging with risk reversals seem to have better performance. In case of puts vanna yields smaller variation than $\zeta_1\zeta_2$ and traditional vega perform slightly better than ζ_1. However for calls this relations are inverted.

The aforementioned asymmetry in the empirical error distribution may come from different sources. First, the performance of the strategies can be biased by behavior of the underlying. During the analyzed time period DAX revealed mainly downward movements, which highly influenced the unhegded portfolios of long barrier options. One observes the 81% of knock-outs in down-and-out put options but only 10% in up-and-out call options. Only 5% of the puts and 39% calls expired in-the-money. The behavior of naked portfolios is summarized in the Table 4.7. From these facts we conclude that the long position in the considered barrier options suffered mainly losses. Therefore, hedging nonsymmetric securities may also leads to nonsymmetric hedging errors.

option	barrier	strike	knock-out	in-the-money
up-and-out call	140%	80%	10%	39%
down-and-out put	80%	110%	81%	5%

Table 4.7: *Information table about the analyzed barrier options. The percentage of are taken as a relation to spot in the first day of the option contract. The percentage of knock-outs refers to the contracts that breached barrier and percentage of in-the-money refers to the options that yielded a positive payoff at expiry.*

Another reason would lead us to the gap risk resulting from the noncontinuous payoffs. If the option knocks out one still owns the hedge portfolio. This gap risk is reflected in the price of the option and hence in all greeks, but besides this it is not considered in the hedging procedure. We do not include any scenario analysis which may take into account the impact of the sudden knock-out and adjust the portfolio accordingly. Therefore when the knock-out occurs the whole portfolio may be subjected to significant loss or gain, because the asset position is suddenly closed. Just before the knock-out the underlying position is boosted up due to the big delta and reaching the barrier level not only annihilates the exotic option but also strongly influence the spot position. If the spot crosses the barrier significantly this effect is even strengthened. The similar effect is for the options near the expiry date, since the absolute values of delta increases as the time to maturity decreases.

To have better overview of this properties we refer once more to Tables 4.8 and 4.9. We report the statistics of the experiment artificially stopped 1 day,

5 days and 25 days before the expiry. The maximum values of each strategy jumps significantly as the standard deviations do. The distribution of the hedging errors become also less skewed and less heavy tailed, which can be observed by the skewness and kurtosis. The differences are not so pronounced in the long horizons as in the final day switching. The location indicators stays, however, rather stable. All these facts highlights the importance of the expiry effect in the final hedge performances. To illustrate it even better we show in Figure 4.28 the standard deviations of hedge errors as a function of the options' life. It is intuitive to expect this function to increase. One may additionally observe significant jump just before the expiry. This figure also illustrates that through the options' life all strategies yield similar standard deviations for put options and the factor hedging brings improvements to the standard methods for call options.

The presented results show how the DSFM can be incorporated in the hedging of financial products. Constructing the sensitivity measures with respect to the statistically motivated risk factors may be an alternative to the traditional methods. Here we do not support the strong arguments for superior performance of this hedging procedure but we just claim that it does the same job. Implementing the trading and risk management systems is a complex task and many different aspects have to be taken into consideration. We show how the surface or curve (e.g. in interest rate management) risk could be factorized and according hedging procedures introduced. It seems to be intuitive in case when the clear interpretation of the factors can be found. In fact the procedure is not much more burdensome. Some aversion may come from the nonparametric character of the functions $\widehat{m}$, however they are easily replaceable by parametric counterparts.

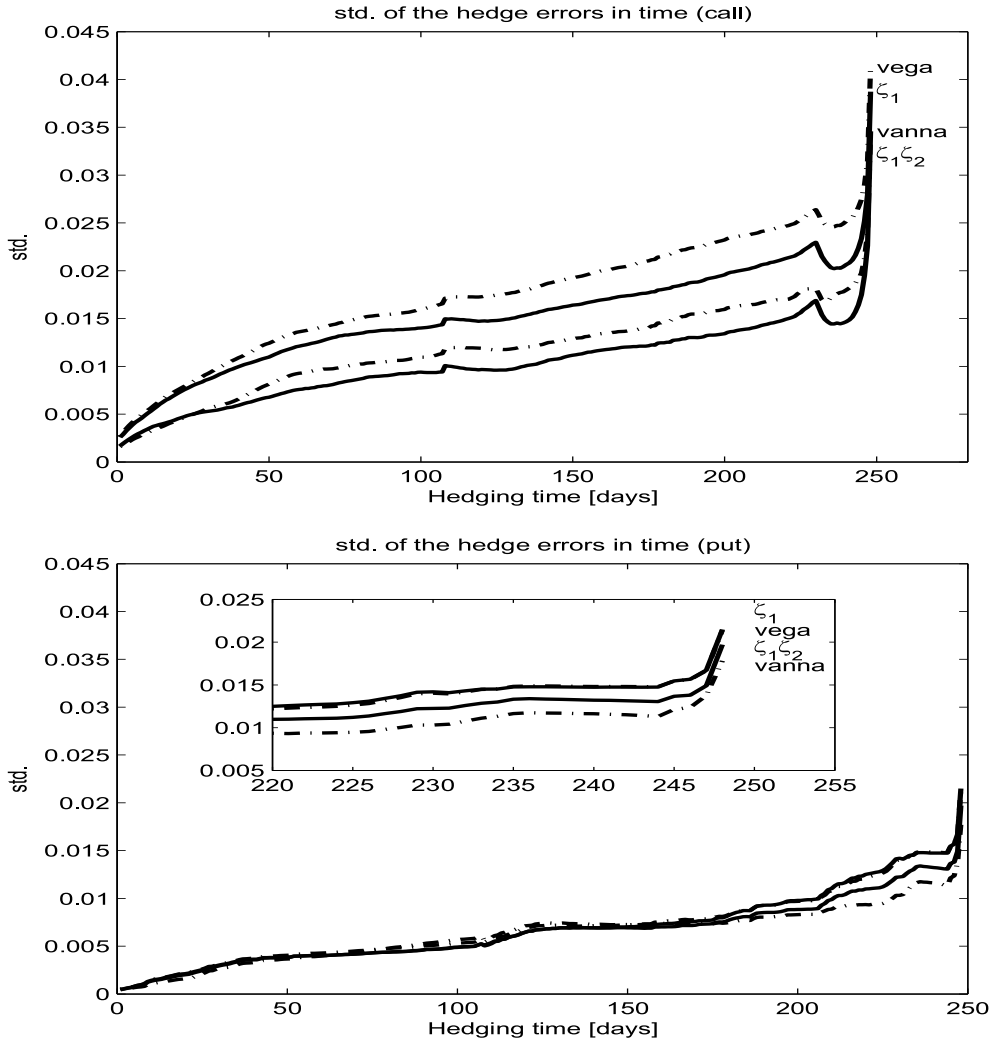

Figure 4.28: *Standard deviations of hedging errors as a function of time from the option issuance. Solid lines stands for the factor hedging methods motivated by the DSFM. Dashed lines represent the vega and vanna hedges. Upper panel: up-and-out call. Lower panel: down-and-out put.*

	days	min	max	mean	median	std.	madev.	skew.	kurt.
vega	0	-0.1038	0.5813	-0.0165	-0.0175	0.0413	0.0209	7.2801	97.60
	1	-0.1038	0.2581	-0.0172	-0.0174	0.0314	0.0199	2.3526	19.32
	5	-0.1037	0.0970	-0.0181	-0.0169	0.0260	0.0183	0.3636	4.91
	25	-0.0827	0.0649	-0.0174	-0.0164	0.0249	0.0178	0.0587	3.74
ζ_1	0	-0.0752	0.5768	-0.0118	-0.0136	0.0387	0.0183	8.4877	119.04
	1	-0.0751	0.2332	-0.0125	-0.0134	0.0279	0.0172	2.8026	22.27
	5	-0.0749	0.0755	-0.0134	-0.0121	0.0216	0.0155	0.2846	4.60
	25	-0.0761	0.0573	-0.0134	-0.0127	0.0215	0.0161	0.0343	3.55
vanna	0	-0.1340	0.5310	-0.0081	-0.0138	0.0345	0.0151	8.6289	124.62
	1	-0.1340	0.1842	-0.0089	-0.0136	0.0239	0.0140	2.1325	17.40
	5	-0.1339	0.0807	-0.0099	-0.0131	0.0187	0.0121	0.2157	9.54
	25	-0.0582	0.0772	-0.0096	-0.0141	0.0173	0.0118	1.3367	6.22
$\zeta_1\zeta_2$	0	-0.0830	0.5684	-0.0066	-0.0119	0.0345	0.0137	10.6470	161.00
	1	-0.0829	0.2091	-0.0073	-0.0117	0.0226	0.0126	4.0718	31.51
	5	-0.0829	0.0710	-0.0083	-0.0113	0.0157	0.0106	1.3447	7.12
	25	-0.0370	0.0629	-0.0086	-0.0118	0.0152	0.0108	1.3559	5.72

Table 4.8: *Descriptive statistics for the hedging strategies of the up-and-out calls* 0, 1, 5 *and* 25 *days before the expiration.*

	days	min	max	mean	median	std.	madev.	skew.	kurt.
vega	0	-0.0264	0.2799	0.0058	-0.0004	0.0213	0.0105	5.4903	51.91
	1	-0.0756	0.1172	0.0050	-0.0004	0.0166	0.0098	2.4531	12.38
	5	-0.0187	0.0882	0.0041	-0.0008	0.0147	0.0090	2.4682	10.92
	25	-0.0186	0.0749	0.0038	-0.0004	0.0124	0.0083	2.0267	8.73
ζ_1	0	-0.0210	0.2808	0.0080	0.0016	0.0214	0.0107	5.4775	51.59
	1	-0.0702	0.1215	0.0072	0.0015	0.0167	0.0100	2.4501	12.14
	5	-0.0137	0.0882	0.0063	0.0013	0.0147	0.0091	2.4112	10.29
	25	-0.0113	0.0798	0.0059	0.0014	0.0127	0.0085	2.0632	8.54
vanna	0	-0.0608	0.2072	0.0022	-0.0016	0.0178	0.0081	5.7326	53.32
	1	-0.0955	0.1309	0.0014	-0.0016	0.0137	0.0074	2.7735	23.76
	5	-0.0323	0.0649	0.0006	-0.0018	0.0114	0.0069	1.9306	9.98
	25	-0.0205	0.0582	0.0004	-0.0016	0.0093	0.0059	2.0273	9.23
$\zeta_1\zeta_2$	0	-0.0332	0.2676	0.0065	0.0008	0.0196	0.0092	6.0258	60.47
	1	-0.0824	0.1146	0.0057	0.0008	0.0149	0.0085	2.4463	14.06
	5	-0.0234	0.0774	0.0048	0.0007	0.0130	0.0079	2.4134	10.42
	25	-0.0121	0.0727	0.0045	0.0010	0.0110	0.0069	2.3842	10.27

Table 4.9: *Descriptive statistics for the hedging strategies of the down-and-out puts* 0, 1, 5 *and* 25 *days before the expiration.*

Bibliography

Ahn, S. K. (1988). Distribution for residual autocovariances in multivariate. *Biometrica*, *75*, 590–593.

Alexander, C. (2001). Principles of the skew. *RISK*, *14*(1), S29–S32.

Amemiya, T. (1971). The Estimation of the Variances in a Variance-Components Model. *International Economic Review*, *12*, 1-13.

Andersen, L. B. G., Andreasen, J., and Eliezer, D. (2002). Static replication of barrier options: Some general results. *Journal of Computational Finance*, *5*(4), 1–25.

Andersen, L. B. G., and Brotherton-Ratcliffe, R. (1997). The equity option volatility smile: An implicit finite-difference approach. *Journal of Computational Finance*, *1*(2), 5–37.

Angelus, A. (2001). Electricity price forecasting in deregulated markets. *The Electricity Journal*, *14*(3), 32–41.

Ayache, E., Henrotte, P., Nassar, S., and Wang, X. (2004). Can anyone solve the smile problem? *Wilmott magazine*, *Jan.*, 78–96.

Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, *25*, 390–397.

Baltagi, B. (2005). *Econometric Analysis of Panel Data*. Chichester: John Wiley & Sons.

Bandettini, P., Wong, E., Hinks, R., Tikofsky, R., and Hyde, J. (1992). Time course of EPI of human brain function during task activation. *Magnetic Resonance in Medicine*, *25*, 390–397.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, *9*, 69–107.

Baune, A., and Sommer, F. (1999). Dynamic cluster analysis of cortical fmri activation. *NeuroImange*, *9*, 477-489.

Bell, A., and Sejnowski, T. (1995). An information-maximization approach to blind separationand blind deconvolution. *Neural Computation*, *7*, 1129-1159.

Benko, M. (2006). *Functional data analysis with application in finance*.

Unpublished doctoral dissertation, Humboldt-Universität zu Berlin, Berlin.

Benko, M., Härdle, W., and Kneip, A. (2008). Common functional principal components. *Annals of Statistics*, *forthcoming*.

Benz, E., and Trück, S. (2006). Modeling the price dynamics of CO_2 emission allowances. *University of Bonn, Working Paper*.

Billingsley, P. (1995). *Probability and measure*. New York: John Wiley & Sons.

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, *34*, 537-541.

Black, F., and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*, 637–654.

Bliss, R. (1997). *Movements in the term structure of interest rates* (Economic Review No. Q IV). Federal Reserve Bank of Atlanta.

Böhringer, C., and Lange, A. (2005). Economic Implications of Alternative Allocation Schemes for Emission Allowances. *Scandinavian Journal of Economics*, *107*(3), 563-581.

Boor, C. de. (2001). *A practical guide to splines*. New York: Springer-Verlag.

Borak, S., Fengler, M. R., and Härdle, W. (2005). DSFM fitting of implied volatility surfaces. In *Conference proceedings of the fifth international conference on intelligent systems design and applications*.

Borak, S., Härdle, W., Mammen, E., and Park, B. (2007). *Time series modelling with semiparametric factor dynamics* (Discussion Paper). Humboldt-Universität zu Berlin: SfB 649.

Borenstein, S., Bushnell, J., and Knittel, C. (1999). Market power in electricity markets: Beyond concentration measures. *The Energy Journal*, *20*, 65–88.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes*. New York: Springer.

Botterud, A., Bhattacharyya, A., and Ilic, M. (2002). Futures and spot prices - an analysis of the Scandinavian electricity market. In *Proceedings of north american power symposium 2002*.

Bouchouev, I., and Isakov, V. (1999). Uniqueness, stability and numerical methods for the inverse problem that arises in financial markets. *Inverse Problems*, *15*, R95–R116.

Breitung, J., and Eickmeier, S. (2005). *Dynamic factor models* (Discussion Paper). Deutsche Bundesbank.

Brüggemann, R., Härdle, W., Mungo, J., and Trenkler, C. (2006). *VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings* (Discussion Paper). Humboldt-Universität zu Berlin: SfB 649.

Brüggemann, R., Lütkepohl, H., and Saikkonen, P. (2006). Residual autocorrelation testing for vector error correction models. *Journal of Econometrics*, *134*, 579–604.

Bunn, D. (2004). Structural and behavioural foundations of competititve electricity prices. In D. Bunn (Ed.), *Modelling prices in competitive electricity markets* (p. 1-17). Wiley.

Burtraw, D. (1996). *Cost savings sans allowance trades? evaluating the so_2 emission trading program to date* (Discussion Paper 95-30-REV). RFF.

Burtraw, D., Palmer, K., Bharvirkar, R., and Paul, A. (2002). The Effect on Asset Values of the Allocation of Carbon Dioxide Emission Allowances. *The Electricity Journal*.

Calhoun, V., Adali, T., Pearlson, G., and Pekar, J. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task related waveforms. *Human Brain Mapping*, *13*, 43-53.

Carr, P., and Chou, A. (1997). Breaking Barriers. *Risk Magazine*, *10*, 139-145.

Carr, P., Ellis, K., and Gupta, V. (1998). Static hedging of exotic options. *Journal of Finance*, *53*(3), 1165-1190.

Carr, P., Geman, H., Madan, D., and Yor, M. (2002). The fine structure of asset returns: An empirical investigation. *Journal of Business*, *75*.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 629–637.

Chen, C., Härdle, W., and Unwin, A. (2007). *Handbook of data visualization*. Berlin: Springer.

Connor, G., Hagmann, M., and Linton, O. (2007). *Efficient semiparametric estimation of the fama-french model and extensions* (Preprint). LSE.

Connor, G., and Linton, O. (2007). Semiparametric estimation of a characteristic-based factor model of stock returns. *Journal of Empirical Finance*, *forthcoming*.

Cont, R., and Fonseca, J. da. (2002). The dynamics of implied volatility surfaces. *Quantitative Finance*, *2*(1), 45–60.

Cont, R., and Tankov, P. (2004). *Financial modelling with jump processes*. London: Chapman & Hall, CRC Press.

Coppersmith, D., and Winograd, S. (1990). Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation*, *9*, 251–280.

Derman, E. (1999). Regimes of volatility. *RISK*, *12*(4), 55-59.

Derman, E., Ergener, D., and Kani, I. (1995). Static options replication. *Journal of Derivatives*, *2*(4), 78–95.

Derman, E., and Kani, I. (1994). Riding on a smile. *RISK*, *7*(2), 32–39.

Detlefsen, K., and Härdle, W. (2006). *Forecasting the term structure for*

variance swaps (Discussion Paper). Humboldt-Universität zu Berlin: SfB 649.

Detlefsen, K., and Härdle, W. (2007). Calibration risk for exotic options. *Journal of Derivatives*, *14*(4), 47-63.

Diebold, F. X., and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, *130*, 337-364.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, *131*, 309-338.

Duann, J., Jung, T., Kuo, W., Yeh, T., Makeig, S., Jen-Chuen, H., et al. (2002). Single-trial variability in event-related BOLD signals. *NeuroImange*, *15*, 823-835.

Dumas, B., Fleming, J., and Whaley, R. E. (1998). Implied volatility functions: Empirical tests. *Journal of Finance*, *80*(6), 2059–2106.

Dupire, B. (1994). Pricing with a smile. *RISK*, *7*(1), 18–20.

Ederington, L., and Guan, W. (2007). Higher order greeks. *Journal of Derivatives*, *14*, 7-34.

Ellerman, A., and Montero, J.-P. (1998). The declining trend in sulfur dioxide emissions: Implications for allowance prices. *Journal of Environmental Economics and Management*, *36*, 26-45.

Engelmann, B., Fengler, M., Nalholm, M., and Schwendner, P. (2007). Static versus Dynamic Hedges: An Empirical Comparison for Barrier Options. *Review of Derivatives Research*, *9*(3), 239-264.

Engelmann, B., Fengler, M., and Schwendner, P. (2006). *Better than its reputation: An empirical hedging analysis of the local volatility model for barrier options* (Working Paper). Available at SRRN.

Fama, E. F., and French, K. R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*, 427–465.

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan, J., Yao, Q., and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. Roy. Statist. Soc. B.*, *65*, 57–80.

Feller, W. (1968). *An introduction to probability theory and its applications*. New York: John Wiley & Sons.

Fengler, M. R. (2005a). *Arbitrage-free smoothing of the implied volatility surface* (Discussion Paper). Humboldt-Universität zu Berlin: SfB 649.

Fengler, M. R. (2005b). *Semiparametric modeling of implied volatility*. Berlin, Heidelberg: Springer-Verlag.

Fengler, M. R., Härdle, W., and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics*, *5*(2), 189-218.

Fengler, M. R., Härdle, W., and Villa, C. (2001). *The dynamics of implied volatilities: A common principle components approach* (Discussion Paper No. 2001-38). Humboldt-Universität zu Berlin: SfB 373.

Figlewski, S., and Green, T. (1999). Market risk and model risk for financial institution writing options. *Journal of Finance*, *54*, 1465-1499.

Filzmoser, P., Baumgartner, R., and Moser, E. (1999). A hierarchical clustering method for analysis functional mr images. *Magnetic Resonance Imaging*, *17*, 817-826.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, *82*, 540-554.

Forni, M., and Lippi, M. (2001). The generalized factor model: representation theory. *Econometric Theory*, *17*, 1113-1141.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W. (2007). *Statistical parametric mapping: The analysis of functional brain images*. London: Academic Press.

Friston, K., Frith, C., Fletcher, P., Liddle, P., and Frackowiak, R. (1996). Functional topography: multidimensional scaling and functional connectivity in the brain. *Cerebral Cortex*, *6*, 156-164.

Friston, K., Frith, C., Liddle, P., and Frackowiak, R. (1993). Functional conectivity: the principal component analysis of large data sets. *Journal of Cerebral Blood Flow & Metabolism*, *13*, 5-14.

Friston, K., Phillips, J., Chawla, D., and Buchel, C. (2000). Nonlinear pca: characterizing interactions between modes of brain activity. *Philosophical Transactions of the Royal Society B:Biological Sciences*, *355*, 135-146.

Friston, K., Poline, J., Holmes, A., Frith, C., and Frackowiak, R. (1996). A multivariate analysis of PET activation studies. *Human Brain Mapping*, *4*, 140-151.

Gao, F., Guan, X., Cao, X., and Papalexopoulos, A. (2000). Forecasting power market clearing price and quantity using a neural network method. In *Proceedings of the ieee power engineering society summer meetting 4* (p. 2183-2188).

Gasser, T., Möcks, R., and Verleger, R. (1983). Selavco: A method to deal with trial-to-trial variability of evoked potential. *Electroencephalography and Clinical Neurophysiology*, *55*, 717–723.

Geman, H., and Roncoroni, A. (2006). Understanding the fine structure of electricity prices. *Journal of Business*, *79*(3), 1225-1262.

Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. Aigner and A. Goldberger (Eds.), *Latent variables in socio-economic models* (p. 365-383). Amsterdam: North Holland.

Giacomini, E., and Härdle, W. (2007). *Statistics of Risk Aversion* (Discussion Paper). Humboldt-Universität zu Berlin: SfB 649.

Glover, G. (1999). Deconvolution of impuls responce in event-related bold fmri. *NeuroImange*, *9*, 416-429.

Green, R., and Newbery, D. (1992). Competition in the British electricity spot market. *Journal of Political Economy*, *100*, 929–953.

Hafner, R. (2004). *Stochastic implied volatility*. Berlin: Springer.

Hafner, R., and Wallmeier, M. (2001). The dynamics of DAX implied volatilities. *International Quarterly Journal of Finance*, *1*(1), 1–27.

Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D. (2002). Managing smile risk. *Wilmott magazine*, *1*, 84–108.

Hallin, M., and Liska, R. (2007). Determining the number of factors in the generalized dynamic factor model. *Journal of the American Statistical Association*, *102*, 603-617.

Hansen, L. H., Nielsen, B., and Nielsen, J. P. (2004). *Two sided analysis of the variance with a latent time series* (Nuffield Colledge Economic Working Paper No. 2004-W25). University of Oxford.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge, UK: Cambridge University Press.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and semiparametric models*. Berlin, Heidelberg: Springer-Verlag.

Härdle, W., and Simar, L. (2003). *Applied multivariate statistical analysis*. Berlin, Heidelberg: Springer-Verlag.

Hastie, T., and Tibshirani, R. (1990). *Generalized additive models*. London: Chapman and Hall.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, *6*, 327–343.

Heston, S., and Nandi, S. (2000). A closed-form garch option valuation model. *Review of Financial Studies*, *13*, 585–625.

Heynen, R. (1994). An empirical investigation of observed smile patterns. *Review of Futures Markets*, *13*, 317–353.

Hinz, J. (2003). Modelling day-ahead electricity prices. *Applied Mathematical Finance*, *10*(3), 149-161.

Hirsa, A., Courtadon, G., and Madan, D. (2003). The effect of model risk on the valuation of barrier options. *Journal of Risk Finance*, *4*, 47–55.

Hobbs, B., Metzler, C., and Pang, J. (2000). Strategic gaming analysis for electric power systems: An MPEC approach. *IEEE Transaction on Power Systems*, *15*, 638–645.

Horowitz, J. (1998). *Semiparametric methods in econometrics* (No. 131). Berlin, Heidelberg: Springer-Verlag.

Hosking, J. R. M. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, *75*, 602–608.

Hosking, J. R. M. (1981). Lagrange-multiplier tests of multivariate time-series models. *Journal of the Royal Statistical Society B*, *43*(2), 219–230.

Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge, UK: Cambridge University Press.

Hull, J., and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, *42*, 281–300.

Hull, J. C., and Suo, W. (2002). A methodology for assessing model risk and its application to the implied volatility function model. *Journal of Financial and Quantitative Analysis*, *37*(2), 297-318.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York: John Wiley & Sons.

ISI. (2003). *Leitfaden und Glossar zum Emissionhandel*. Fraunhofer Institut für Systemtechnik und Innovationsforschung (ISI).

Johnson, R. A., and Wichern, D. W. (1998). *Applied multivariate statistical analysis* (4 ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.

Karatzas, I. (1997). *Lectures on the mathematics of finance* (Vol. 8). Providence, Rhode Island: American Mathematical Society.

Kauermann, G. (2000). Modeling longitudinal data with ordinal response by varying coefficients. *Biometrics*, *56*(3), 1692–1698.

Kirschen, D. (2003). Demand-side view of electricity markets. *IEEE Transaction on Power Systems*, *18*(2), 520–527.

Koekebakker, S., and Ollmar, F. (2005). Forward Curve Dynamics in the Nordic Electricity Market. *Managerial Finance*, *31*(6), 73-94.

Krivobokova, T., Kauermann, G., and Archontakis, T. (2006). Estimating the term structure of interest rates using penalized splines. *Statistical Papers*, *47*(3), 443-459.

Kwong, K., Belliveau, J., Chesler, D., Goldberg, I., and R.M., W. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of USA*, *89*, 5675-5679.

Latané, H. A., and Rendelman, J. (1976). Standard deviations of stock price ratios implied in option prices. *Journal of Finance*, *31*, 369–381.

Lee, R. D., and Carter, L. (1992). Modeling and forecasting the time series of u.s. mortality. *Journal of the American Statistical Association*, *87*(419), 659-671.

Liao, R., , M. , McKeown, and Krolik, J. (2006). Isolation and minimization

of head motion-induced signal variations in fmri data using independent component analysis. *Magnetic Resonance in Medicine*, *55*, 1396–1413.

Linton, O., and Nielsen, J. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrica*, *82*, 93–101.

Logothetis, N., and Wandell, B. (2004). Interpreting the bold signal. *Annual Review Physiology*, *66*, 735-769.

Longstaff, F., and Wang, A. (2004). Electricity foraward prices: A high-frequency empirical analysis. *Journal of Finance*, *59*(4), 1877-1900.

Lukic, A., Wernick, M., Hansen, L., Anderson, J., and Strother, S. (2002). A spatially robust ica algorithm for multiple fmri data sets. In *Ieee international symposium on biomedical imaging, proceedings* (p. 839-842).

Lütkepohl, H. (1993). *Intorduction to multiple time series analysis*. Berlin, Heidelberg: Springer-Verlag.

Lütkepohl, H. (2004). *Intorduction to multiple time series analysis*. Cambridge: Cambridge University Press.

Madan, D., and Seneta, E. (1990). The variance gamma process for share market returns. *Journal of Business*, *63*, 511-524.

Maeda, A. (2001). *Domestic greenhouse gas emissions trading markets: Forward pricing and banking impacts* (Working Paper No. IR-01-048). International Institute for Applied Systems Analysis, Laxenburg.

Mammen, E., Linton, O., and Nielsen, J. (1999). The existence and asymptotic properties of backfitting projection algorithm under weak conditions. *Annals of Statistics*, *27*, 1443-1490.

Martinussen, T., and Scheike, T. (2000). A nonparametric dynamic additive regression model for longitudinal data. *Annals of Statistics*, *28*(4), 1000–1025.

McIntosh, A., Bookstein, F., Haxby, J., and Grady, C. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImange*, *3*, 143-157.

McKeown, M., Makeig, S., Brown, G., Jung, T., Kindermann, S., Bell, A., et al. (1998). Analysis of fmri data by blind separation into independent spatial components. *Human Brain Mapping*, *6*, 160-188.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, *4*(Spring), 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125-144.

Meyer, F., and Chinrungrueng, J. (2005). Local clustering of fmri time series in the frequency domain. *Medical Image Analysis*, *9*, 51-68.

Miezin, F., Macotta, L., Ollinger, J., Petersen, S., and Buckner, R. (2000).

Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImange*, *11*, 735-759.

Molgedey, L., and Galic, E. (2001). Extracting factors for interest rate scenarios. *European Physical Journal B*, *20*(4), 517–522.

Musiela, M., and Rutkowski, M. (1997). *Martingale methods for financial modelling*. Berlin, Heidelberg: Springer-Verlag.

Nalholm, M., and Poulsen, R. (2006a). Static hedging and model risk for barrier options. *Journal of Future Markets*, *26*, 449-463.

Nalholm, M., and Poulsen, R. (2006b). Static hedging of barrier options under general asset dynamics: Unification and application. *Journal of Derivatives*, *13*, 46-60.

Nelson, C. R., and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, *60*, 473–489.

Nerlove, M. (1971). A note on error components models. *Econometica*, *39*, 383-396.

Nogales, F., Contreras, J., Conejo, A., and Espinola, R. (2002). Forecasting next day electricity prices by time series models. *IEEE Transaction on Power Systems*, *17*, 342–348.

Nowicka-Zagrajek, J., and Weron, R. (2002). Modelling electricity loads in California: ARMA models with hyperbolic noise. *Signal Processing*, *82*, 1903–1915.

Ogawa, S., and Lee, T. (1990). Magnetic resonance imaging of blood vessel at high fields: in vivo and in vitro measurements and image simulation. *Magnetic Resonance in Medicine*, *16*, 9-18.

Ogawa, S., Lee, T., Kay, A., and Tank, D. (1990). Intrisic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of USA*, *87*, 9868-9872.

Ogawa, S., Lee, T., Nayak, A., and Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, *14*, 68-78.

Ogawa, S., Tank, D., Menon, R., Ellermann, J., and Kim, S. (1992). Intrisic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of USA*, *89*, 5951-5955.

Paolella, M., and Taschini, L. (2006). *An econometric analysis of emission trading allowances* (Working Paper). Swiss Banking Institute.

Peña, D., and Box, E. P. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, *82*, 836–843.

Petersson, K. M., Nichols, T. E., Poline, J. B., and Holmes, A. P. (1999). Sta-

tistical limitations in functional neuroimaging. I. Non-inferencial methods and statistical models. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *354*(1387), 1239–1260.

Purdon, P., Solo, V., Weisskoff, R., and Brown, E. (2001). Locally regularized spatiotemporal modeling and model comparison for functional mri. *NeuroImange*, *14*, 912-923.

Qui, S., and Lane, T. (2006). *Parallel Kernel Computation for High Dimensional Data and Its Application to fMRI Image Classification* (Technical Report Number TR-CS-2004-12). University of New Mexico.

Ramsay, J. O., and Silverman, B. W. (1997). *Functional data analysis*. Berlin, Heidelberg: Springer-Verlag.

Rebonato, R. (1998). *Interest-rate option models: Understanding, analyzing and using models for exotic interest-rate options* (2nd ed.). John Wiley & Son Ltd.

Rezek, J. (1999). *Shadow prices of sulfur dioxide allowances in Phase I Electric Utilities.* (Annual meeting of the American Agricultural Economics Association)

Rosenberg, J. (2000). Implied volatility functions: A reprise. *Journal of Derivatives*, *7*, 51–64.

Rudin, W. (1991). *Functional analysis*. New York: McGraw-Hill.

Ruppert, D., Wand, M., and Carroll, R. (2003). *Semiparametric regression*. New York: Cambridge University Press.

Samuelson, P. A. (1965). Rational theory of warrant prices. *Indust. Manag. Rev.*, *6*, 13-31.

Schennach, S. (2000). The economics of pollution permit banking in the context of Title IV of the 1990 Clean Air Act Amendment. *Journal of Environmental Economics and Management*, *40*, 189–210.

Schleich, J., Ehrhart, K., Hoppe, C., and Seifert, S. (2006). Banning Banking in EU Emissions Trading. *Energy Policy*, *34*(1), 112-120.

Schöbel, R., and Zhu, J. (1999). Stochastic Volatility With an Ornstein-Uhlenbeck Process: An Extension. *European Finance Review*, *3*(1), 23-46.

Schoutens, W., Simons, E., and Tistaert. (2004). A perfect calibration! now what? *Wilmott Magazine*.

Shao, J. (2003). *Mathematical statistics*. New York: Springer-Verlag.

Skiadopoulos, G., Hodges, S., and Clewlow, L. (1999). The dynamics of the S&P 500 implied volatility surface. *Review of Derivatives Research*, *3*, 263–282.

Stein, E. M., and Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies*, *4*, 727–752.

Stock, J. H., and Watson, M. W. (2005, July). *Implications of dynamic factor models for var analysis* (NBER Working Papers No. 11467). National Bureau of Economic Research, Inc. (available at http://ideas.repec.org/p/nbr/nberwo/11467.html)

Stoft, S. (2002). *Power Systems Economics: Designing Markets for Elecricity*. Wiley-IEEE Press.

Svensén, M., Kruggel, F., and Cramon, D. von. (2000). Probabilistic modeling of single-trial fmri data. *IEEE Transactions on Medical Imaging*, *19*, 25-35.

Swamy, P., and Arora, S. (1972). The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models. *Econometica*, *40*, 261-275.

Szkuta, B., Sanabria, L., and Dillon, T. (1999). Electricity price short-term forecasting using artificial neural networks. *IEEE Transaction on Power Systems*, *14*(3), 851–857.

Tabelow, K., Polzehl, J., Voss, H., and Spokoiny, V. (2006). Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImange*, *33*, 55-62.

Thirion, B., and Faugeras, O. (2003). Dynamic components analysis of fmri data through kernel pca. *NeroImage*, *20*, 34-49.

Tompkins, R. (2002). Static versus dynamic hedging of exotic option: An evaluation of hedge performance via simulation. *The Journal of Risk Finance*, *3*, 6–34.

Trück, S., Borak, S., Härdle, W., and Weron, R. (2006). *Convenience yields for CO_2 emission allowance futures contracts* (Discussion Paper No. 2006-076). Humboldt-Universität zu Berlin: SfB 649.

Uhrig-Homburg, M., and Wagner, M. (2006). Success chances and optimal design of derivatives on co_2 emission certificates. *Working Paper, University of Karlsruhe*.

Ventosa, M., Baillo, A., Ramos, A., and Rivier, M. (2005). Electricity market modelling trends. *Energy Policy*, *33*(7), 897-913.

Wallace, T., and Hussain, A. (1969). The Use of Error Components Models in Combining Cross Section with Time Series Data. *Econometica*, *39*, 55-72.

Weron, R. (2006). *Modelling and Forecasting Electricity Loads and Prices: A Satitstical Approach*. Chichester: John Wiley & Sons.

Wood, A., and Wollenberg, B. (2006). *Power Generation, Operation and Control*. New York: Wiley.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., et al. (2002). A general statistical analysis for fmri data. *NeuroImange*, *15*, 1-15.

Wu, H., and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Hoboken, New Jersey: John Wiley & Son Ltd.

Yang, L., Park, B. U., Xue, L., and Härdle, W. (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *Journal of the American Statistical Association*, *101*, 1212–1227.

Zhang, B., and Dong, Z. (2001). An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research*, *59*, 121-129.

Zhu, Y., and Avellaneda, M. (1997). An E-ARCH model for the term-structure of implied volatility of FX options. *Applied Mathematical Finance*, *4*, 81–100.

Selbständigkeitserklärung

Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.